

Nature in 2003

This year sees *Nature* getting up close and personal with researchers, thanks to a new series and a territorial expansion. Another important innovation is a policy that allows authors to retain copyright.

One of the world's longest-running radio series is the BBC's *Desert Island Discs*, in which celebrities of the day are invited to select the eight recordings that they'd prefer to be marooned with, as well as talk about their favourite subject: themselves. Not immune to the appeal of this format, we are this week launching our own interview series, *Lifelines* (see page 23), in which working scientists (not necessarily celebrities) are provoked into taking a sidelong, irreverent and sometimes profound look at their own lives and tastes. On the grounds that those who do not seek public office are the best qualified for the job, unsolicited contributions to this section will not be considered. (For this we can cite some kind of precedent. Herbert Morrison, a British cabinet minister of an earlier generation, desired nothing more than to appear on *Desert Island Discs*. His desire was so great that he carried a list of his eight favourite recordings in his pocketbook, just in case he got the call. He never did.)

More developments in content are on their way. Later this month, readers will encounter an enhancement of our News and Views section, intended to increase its diversity. And later in the year there will be a new series devoted to the influences of science on the arts.

Another development is geographic. *Nature* already has editorial staff in London, Munich, Paris, Washington DC, San Francisco, San Diego and Tokyo. But if one considers the world's outstanding centres of scientific activity, there is a glaring omission in that list — a city which, despite recent stockmarket upheavals, is experiencing a series of new activities in traditional university departments, multidisciplinary research centres, technological spin-offs and industrial research laboratories. Accordingly, *Nature* and its family of research journals will be represented in Boston, with the opening of a new office there.

These editorial bases, coupled with the New York base of six of our eight sister research journals (one of which, *Nature Genetics*, has been celebrating its tenth anniversary year), give us the benefits of proximity and, through closer engagement, a greater sense of kinship in pursuing the interests and needs of researchers, whether as authors or readers.

New services to authors will be announced before long, but perhaps the most significant development on this front is that *Nature* and all other journals published by the Nature Publishing Group (NPG) have introduced a new policy in relation to copyright. No longer do we require authors of papers to sign away their copyright. Instead, we now ask authors to grant NPG the exclusive licence to publish the paper in all media throughout the world, to translate it into other languages, and to adapt it or license it to others. (If all co-authors are US government employees, slightly different arrangements apply.)

Ownership of copyright remains with the authors. Provided that, when reproducing their contribution or extracts from it, the authors acknowledge the original publication in *Nature* or other NPG journal, they may reproduce the paper in any printed volume of which they are the authors. Furthermore, they and any academic institution where they work at the time may reproduce the paper without payment for the purpose of course teaching.

Authors may also post a copy of their paper on their own website once the printed edition has been published, provided that they also provide a link from the contribution to *Nature's* website. "Their own" refers to any site devoted to them, whether owned by them or by a not-for-profit employer. However, it does not mean open archival websites, such as those that host collections of articles by an institution's researchers, which would amount to a breach of our licence.

This policy is being applied retrospectively. Hundreds of thousands of scientists are authors of papers covered by copyright agreements that are still in force, and we cannot renegotiate every agreement. But we are happy to extend to all past authors the rights laid out in the new licence agreements: to re-use the papers in any printed volume of which they are an author; to post a PDF copy on their own (not-for-profit) website; to copy (and for their institutions to copy) their papers for use in coursework teaching; and to re-use figures and tables. For the exact terms and conditions, please see a copy of the licence agreement at <http://npg.nature.com/authornews>. ■

Overseas abuse of China's development

China must do more to protect the integrity of its policies that encourage greater participation by Chinese researchers overseas.

For every billion in funds that the Chinese government pours into research, there are probably a hundred overseas researchers ready to return to take advantage of its generosity. But the investment boom aimed at strengthening China's science greatly outweighs the development of administrative infrastructure needed to monitor the system.

A petition has brought the problem to the attention of the Chinese scientific community. Critics say that overseas Chinese researchers are rapaciously gathering grants in China while reneging on agreements to spend most of their time overseeing the research projects there (see page 23). The critics, most of whom gave up positions abroad to work full-time in China, rightly want the government to act.

This will require the government to do something it is not used to doing: establish a greater rapport with a wider base of its researchers.

In other words, it needs to complement its notoriously top-down approach to scientific investment with a bottom-up element, whereby a wide range of China's researchers can give feedback to the government, not only on issues of misconduct and cheating, but also on investment priorities.

Without determined action, some researchers will continue to get big grants and big laboratories without putting in the time. The research institutions or universities that house them stand to gain from the grants' overhead contributions, and might thus tolerate abuse. But Chinese science will lose out. It will not only lose the investment, through inefficient use, but it could also lose the faith of researchers as it blurs the line between those who are running back and forth between two countries honestly trying to help China, often with little recompense, and those who are merely out for gain. ■





Bacterial bother
US researchers
wrestle with legality
of pathogen storage
p4



All at sea
Budget cuts leave
ocean institute
fearful for the future
p5



Tuning in
Committee takes
control of Puerto
Rican telescope
p6



Net loss
Europe's fishermen
up in arms over
scientific quotas
p6

Petition calls for clampdown on absentee Chinese researchers

David Cyranoski

A group of Chinese scientists is calling for a crackdown on colleagues who, they allege, receive the most sought-after grants but fail to devote enough time to research within China.

China has watched many of its best young researchers head off to the United States and elsewhere over the past decade, and is now trying lure hundreds of them back, often on generous terms. But as the petition shows, the effort is creating tensions in China's rapidly expanding scientific community.

Some grantees are using the money to set up laboratories in China but then spend most of their time abroad, according to a letter to agency directors signed by 26 researchers from the Chinese Academy of Sciences (CAS) and various universities and research institutions. The letter says that the practice defies the terms of the grants, which require recipients to spend six or nine months in China.

"I am strongly against those who cheat on the working time required by positions or funding in China," says Zhuan Zhou of the CAS Institute of Neuroscience in Shanghai, who signed and helped to circulate the letter.

The petition was sent to the heads of the CAS, the science and technology ministry, the education ministry and the National Nat-



Fair play: Shigang He wants to see grantees carry out more of their research in China.

ural Science Foundation of China (NNSFC) on 19 December.

The letter's author, Shigang He, also of the Institute of Neuroscience, claims that at least a dozen people in the life sciences alone are wasting China's resources. These researchers do most of their important work overseas, leaving largely unsupervised labs in China where graduate students or assistant

researchers do not know how to use the equipment, he says. "The money gets taken away from people who are really qualified and is given to people who are not," He says.

Some heads of institutes within the CAS are already taking measures in response to the researchers' concerns. Yongbiao Xue, associate director of the CAS Institute of Genetics and Developmental Biology in Beijing, says that researchers who fail to meet the time requirements at his institution risk losing their lab.

The issue is of particular concern because grants from several government sources often end up in the same hands, as success with one grant application often paves the way to success with others, says He.

The funding agencies should monitor the use of the grants more closely by checking passports, He suggests, or requiring a written notice from the researcher's previous overseas institution stating it is aware that the researcher is taking on another full-time post. "The rule is there," he says. "They just have to enforce it."

But Chen Jia'er, president of the NNSFC, says that the problem is not so severe, noting that its grants involve strict evaluations and progress reports. But he acknowledges that "the issue raised is something we should pay more attention to". ■

Human cloning claim sparks fear of Senate backlash

Colin Macilwain, Washington

Concern is mounting among US biologists that unconfirmed reports of the birth of cloned babies could force through legislation that will outlaw any cloning of human embryos, even for research purposes.

Senator Sam Brownback (Republican, Kansas) is set to reintroduce legislation in the next few weeks that would enact such a ban. In the past, Brownback has been unable to garner the 60 Senate votes that are needed to pass his legislation, in the face of trenchant opposition from researchers.

On 27 December, however, Clonaid, a

company controlled by the Raelian sect, a religious group that believes in extraterrestrials, told a disbelieving world that a cloned baby had been born at an undisclosed location. The claim — for which no evidence has yet been presented — is likely to be the first of several over the next few weeks from various doctors around the world.

Whether genuine or not, the claim may serve to build support for Brownback's legislation. Importantly, his proposal has the backing of new Senate Republican leader Bill Frist (Tennessee), a heart surgeon who normally supports biomedical research.

Frist was elected on 23 December after Senator Trent Lott (Republican, Mississippi) was forced to resign.

Brownback's bill is also backed by President George Bush, and is similar to a measure that has already been passed by the House of Representatives. Besides making reproductive cloning illegal, it would also outlaw 'therapeutic cloning' of embryos to produce stem cells for research purposes. A rival measure, that would allow therapeutic cloning while banning the cloning of babies, has the support of scientific organizations but lacks the votes to be passed in either the House or the Senate, observers say. ■

Law sends laboratories into pathogen panic

Erika Check, Washington

When unknown parties mailed anthrax spores to several US addresses in the autumn of 2001, plant researchers at a herbarium at Harvard University began to get nervous.

For years, the researchers had stored a set of innocuous-looking brown envelopes that contained samples of anthrax. Within weeks of the attacks, President George W. Bush had signed a law called the USA Patriot Act, under which possession of anthrax without a "bona fide research justification" became a criminal offence. The Harvard researchers soon found themselves facing a tricky dilemma—how to balance their hoarding instincts against the new demands of homeland security.

As thousands of US biologists face up to the same problem, some scientific leaders are concerned that researchers are dumping valuable samples to avoid trouble with the law.

For example, Ron Atlas, president of the American Society for Microbiology, is alarmed by the prosecution last July of a University of Connecticut graduate student who kept anthrax in his freezer. Atlas says these old microbes could hold useful information. "We are really in a delicate balance as to whether individuals will hold on to their cultures or whether they'll feel endangered by the USA Patriot Act," he says. He also warns that those who clear out their freezers may have problems restocking them because of new restrictions

on the movement of pathogens between labs.

Some institutions — such as Iowa State University at Ames, which destroyed its entire archive of anthrax samples in October 2001 — have ordered mass clear-outs of materials. But individual researchers have also taken it upon themselves to dump potentially dangerous microbes. John Collier, a Harvard microbiologist who has long worked on the anthrax toxin, got rid of his samples of the bacteria late in 2001. "I wanted to be able to tell the world we didn't have *Bacillus anthracis*," Collier says.

The issue has now attracted the White House's attention — in part because archived samples could prove useful in criminal investigations. Kathryn Harrington, a spokeswoman for the administration's Office of Science and Technology Policy, says that the office is "aware of the destruction of select agents and is concerned". She adds that the office is trying to "encourage researchers to transfer materials to a secure facility, rather than destroy them".

This is exactly what the Harvard plant scientists did, after consulting their colleague molecular biologist Matthew Meselson. He contacted Paul Keim, a microbiologist at Northern Arizona University in Flagstaff who has spent his career studying anthrax. Keim persuaded the Harvard scientists to send him their envelopes, which are thought to contain anthrax taken from the blood of a cow in 1883. "We were interested in these for basic



Bottling it: biologists are jettisoning their collections to avoid falling foul of legislation.

G. TOMPKINSON/SPL

pathogen-evolution studies," Keim says. "But now they're crucial for fighting bioterrorism."

Keim also argues that the federal government should use some of its new bioterrorism funds to solve the problem once and for all by creating a central repository for pathogens. "There's a big problem with saving these collections and a big problem with getting access to them," Keim says. "What's the point of putting \$1.7 billion into the research if nobody can get hold of the strains?" ■

Biotech critic tries to sew up research on chimaeras

Erika Check, Washington

Scientists who are seeking to meld human embryonic stem cells with mouse embryos have been warned that they could be sued if they pursue the idea.

The 'chimaeric' embryos would be used to test the stem cells' ability to divide into cells with different functions (see *Nature* 420, 255; 2002). But Jeremy Rifkin, an

economist and well-known critic of the biotechnology industry, has told researchers to abandon their plans, claiming that he is about to win a wide-ranging patent on human-animal chimaeras.

"They're saying they cannot take advantage of therapeutic cloning with stem cells unless they place them in an animal model," Rifkin says. "And we're saying we control that."

Rifkin and Stuart Newman, a cell biologist at New York Medical College, applied for the patent in 1997. So far, examiners at the US Patent and Trademark Office have said three times that the application should be turned down — but it remains under review. Rifkin claims that he will prevail in a court appeal even if the patent office denies his claims.

Rifkin's lawyers have sent letters asserting the claims to prominent researchers in the field, including Ali Brivanlou, a developmental biologist at Rockefeller University in New York, Austin Smith of the University of Edinburgh, UK, and James Thomson of the University of Wisconsin at Madison.

But Brivanlou, one of the researchers working on a discussion paper considering the production of chimaeric embryos, says he is undeterred by the letter. "This certainly isn't going to stop me from doing anything," he says. "I'm not taking it seriously."

Brivanlou says that he is highly sceptical of Rifkin's warnings. He points out that Rifkin has not been awarded a patent and that Rifkin and Newman have not done the experiments described in their patent application as proof that it is possible to make a chimaeric embryo.

Rifkin and his lawyer contend that they don't have to make a chimaera to win a patent on it. "There is no rule, regulation, case law or statute of which I'm aware that requires the inventor to practise his or her invention," says Patrick Coyne, Rifkin's lawyer at the Washington firm Collier Shannon Scott.

A lawyer not associated with the case says that although this is technically correct, courts have recently asked for proof that biotechnology inventions actually work before granting patents on them. ■



Bleak outlook for universities as state budget deficits bite

Rex Dalton, San Diego

Public universities in the United States are preparing for severe cuts in some of their research programmes this year, as state legislatures respond to burgeoning budget deficits.

Some institutions — such as the University of California (UC) system, which is already losing \$80 million from its \$320 million of annual state-funded research — are even now reducing their departmental budgets, axing faculty positions and urgently seeking new sources of income.

At the University of Texas at Austin, administrators are pushing for a new state law to let them raise tuition fees. The University of Colorado at Boulder will cut 33 vacant academic posts as part of a plan to save \$11 million this year. And the University of Arizona in Tucson may urge professors to try to cover more of their salaries from granting bodies — although most federal agencies prohibit this.

Many state governments begin their legislative processes this month and see cuts in the universities as one way of balancing budgets that have been thrown into deficit by sharply declining tax revenues. Unlike the federal government, most states are legally required to balance their budgets and have little choice but to cut budgets when their revenues decline.

Public universities that receive state funds for research find themselves in an unusual position: while federal research grants flow steadily from the National Institutes of Health and the National Science Foundation, state funds that support their operations and some types of research are likely to be cut.

This situation creates startling inequities at universities, with medical-school research programmes, for example, continuing to expand, while sharp cuts hit other research activities. Such circumstances are already arising on several of the nine UC campuses, which annually receive a total of about \$2.5 billion in research funds from all sources.

At UC San Diego's Scripps Institution of Oceanography in La Jolla, for instance, state budget cuts of \$1.3 million in recent months are threatening the maintenance of historic collections of natural objects, as well as operating funds for the research vessel *Revelle*. "We feel like sacrificial lambs," says Phil Hastings, Scripps' curator of fishes.

Anticipating deeper cuts, Scripps administrators are scrambling to find alternative financing to support their \$148-million annual budget, and are even talking of a 'doomsday scenario' to dissolve the 100-year-



Out on a limb: budget cuts threaten collections and operating funds at the Scripps Institution.

old institution should the state further reduce its \$10-million annual research contribution. "This is the worst since the Depression," says Scripps' deputy administrator Tom Collins.

At UC Davis' College of Agricultural and Environmental Sciences, meanwhile, state cuts of \$3.8 million are hitting the Agriculture Experiment Station, which conducts statewide crop research in one of the world's biggest food-producing regions. About 25 of the 450 faculty positions may be axed.

And biomedical research programmes are not immune, officials say. UC San Francisco was told last month to cut \$10 million in substance-abuse research, and UC Los Angeles' Brain Injury Research Center, which funds seed grants at all UC campuses, must halve its \$5-million annual budget.

With the announcement in late December that the state of California will face a budget deficit of \$34.8 billion in the fiscal year that begins in July, university officials and academics are looking with some foreboding to 10 January, when governor Gray Davis will release his budget proposal.

Almost every public university system in the country has a similarly bleak outlook. University of Arizona president Peter Likins, for example, says that the state's board of regents, which also governs Arizona State and Northern Arizona universities, has said that it will take "a radically different perspective" and will enact unspecified new policies to "address enormous financial challenges". ■

Joint European plan will tackle Africa's killer diseases

Natasha McDowell, London

The New Year is expected to see the birth of an innovative European Union venture that will support clinical trials of treatments for three of Africa's most devastating diseases: AIDS, tuberculosis and malaria.

European governments and institutions are likely to commit E600 million (US\$620 million) to the venture over five years. Yet some observers are worried that the planned European-Developing Countries Clinical Trials Programme (EDCTP) could stumble in the face of resistance from established national programmes and confusion about its innovative, pan-European approach to research support.

Medicines that show early promise against the three diseases in small-scale clinical trials often progress no further, mainly because poor countries lack the money and infrastructure for larger trials, the EDCTP's advocates say.

However, when the programme was first conceived in 2001, some developing countries were suspicious of Europe's proposed dominant role. Now, organizers say, it will function as an equal partnership between the European Union and developing countries.

The venture is thought to be the first attempt to create a joint research programme between the European Union, its member states and other European sponsors. "The hope is that a joint research programme will give the EDCTP independence from the usual European bureaucracy," says Brian Greenwood, a malaria specialist at the London School of Hygiene and Tropical Medicine.

Some have voiced concern that effective European coordination may end up as a higher priority than the actual development of drugs or vaccines. But if the venture is run by an independent board of scientific experts, as planned, interference by Commission officials will hopefully be avoided, says Greenwood.

The European Parliament is expected to give final approval for funding of the venture by March, and money could begin to flow by the autumn. One-third of it will come from existing, national clinical-research projects, one-third from the European Union and the remainder from research foundations and the pharmaceutical industry. The scientists appointed to the independent board will meet this spring to determine the programme's structure. ■

Biotech joint venture flounders in wake of Syngenta's silence

San Francisco A partnership between biotechnology firm Syngenta and the University of California, Berkeley's plant and microbial biology department may be about to break up.

The university receives US\$5 million a year under the arrangement, which allows Syngenta to see research results before they are published, to request that the university file specific patents, and to be offered first options on licensing agreements. The five-year deal expires this November, but the company allowed last November's renewal deadline to pass without contact. The agreement had been criticized by some faculty members, who said that the academic department was being sold to a private company.

A spokesman at Syngenta's headquarters in Basel, Switzerland, says the company has been focusing on the restructuring of its own plant-genomics research, including last month's decision to close its Torrey Mesa Research Institute in San Diego, and that a final decision on the Berkeley collaboration will be made this year.

Shake up for management at Arecibo radio telescope

Washington Cornell University has set up a special oversight committee for the Arecibo radio observatory in Puerto Rico, which it manages for the National Science Foundation.

The four-person Cornell NAIC (National Astronomy and Ionosphere Center) Oversight Committee will be chaired by Martha Haynes, a Cornell astronomer. Its establishment coincides with the departure of NAIC's director, Paul Goldsmith, and is likely to address long-standing concerns about the management of the world's largest single-dish radio telescope (see *Nature* 384, 505; 1996). During the past year, the US Congress has been pressing the National Science Foundation to tighten its control of major facilities such as Arecibo.



Dish of the day: a taskforce is set to address concerns at Puerto Rico's top telescope.

One of the committee's first tasks, says Haynes, will be to find a new director for the NAIC. The Arecibo observatory is currently adding instruments to conduct astronomical surveys and link to other radio telescopes, and may soon seek support for a new initiative to study the upper atmosphere. A \$27-million upgrade of the observatory, completed in 1997, was marred by schedule and budget overruns, as well as legal disputes with the main contractor, COMSAT RSI, over who should pay the additional costs.

More papers fall foul of Schön inquiry

Washington Last year's highest-profile case of scientific misconduct — the affair of Bell Laboratories physicist Jan Hendrik Schön — has led to the retraction of another six papers.

A Bell Labs report, released last October, found Schön guilty of 16 different cases of fabricating or falsifying data (see *Nature* 419, 419–421; 2002). The report prompted the American Physical Society (APS) to examine all papers in its journals that included Schön as an author, and it announced on 20 December that six publications are to be withdrawn. Four of these were not covered by the Bell Labs report, but have been retracted with the consent of all the authors but Schön.

The APS declined to discuss details of the retraction, saying only that the authors felt “some measure of discomfort” with the papers. Two other papers considered for retraction — including one written solely by Schön — remain in the literature at the request of the authors.

Free-access group secures deal to publish journals

San Francisco The Public Library of Science (PLOS), which in 2001 urged researchers to boycott journals that do not provide free access to papers, has received a US\$9-million grant to help start its own publishing operation.

The group, which campaigns for free access to medical and scientific literature, says that the five-year grant from the San Francisco-based Gordon and Betty Moore Foundation, announced on 17 December, will enable it to begin publishing *PLOS Biology* and *PLOS Medicine* later this year. Online access to both journals will be free, with peer review and other administrative expenditure covered by charging authors around \$1,500 per paper. The PLOS says that print copies will be available at cost.

The move is the latest step in a campaign by the PLOS, which includes several prominent US biologists. Last year the group asked researchers to sign a letter pledging to boycott journals that did not provide free access to material within six months of publication.

♦ www.publiclibraryofscience.org



The chips are down: restriction on fish catches has left many in the industry fearful for their jobs.

Fishermen battered by scale of fish quota

Brussels European fishermen are claiming that the new catch quotas agreed by the European Union (EU) will destroy some fishing communities. Yet scientists advising the EU say the measures may not be enough to avoid the complete collapse of some fish stocks.

The agreement, reached by fisheries ministers on 20 December after a week-long negotiating marathon in Brussels, will see annual catch quotas replaced by science-based management plans. The plan will also mean the end of state subsidies for building and modernizing fishing vessels by 2004. Catch quotas for North Sea cod will be reduced by 45% from February.

But although the cut in cod allowances will mean job losses in the fishing industry, it falls short of the moratorium recommended by the Copenhagen-based International Council for the Exploration of the Sea, which gives scientific advice to fisheries authorities (see *Nature* 419, 866; 2002). “We told policymakers precisely what is going on,” says Hans Lassen, a marine biologist at the council. “Unfortunately, when the advice is out our role has ended.”

Belarus seeks recompense for poached scientists

Minsk The president of the National Academy of Sciences of Belarus says he is tired of losing home-grown talent to well-paid positions in Western organizations. So Mikhail Myasnikov is proposing an innovative solution — asking for compensation from the countries that lure away his colleagues.

Under a proposal being developed at the academy, Western companies and the countries in which they operate could be asked to pay up to US\$600,000 if they recruit a Belarus-educated academic. Myasnikov says that this is the cost to the government of educating a scientist to doctorate level.

The average scientist earns around US\$120 per month in Belarus, and about 70 emigrate annually. The academy says it will not attempt to enforce the plan, but hopes to use it to stimulate discussion at international organizations such as the United Nations.

The law of the jungle

Forest mammals are food for people across Asia, Africa and South America. But with many species disappearing fast, can hunting be made sustainable? John Whitfield talks to the ecologists trying to balance supply and demand.

In west and central Africa, it's estimated that one million tonnes of forest animals are killed for meat each year — equivalent to a daily quarter-pound burger for each of the forests' 30 million people. Road building and the spread of shotguns and wire snares have made hunting a more serious threat to forest wildlife even than deforestation. In 2000, Miss Waldron's red colobus monkey (*Procolobus badius waldroni*), once resident in Ghana and the Ivory Coast, was hunted to extinction. And it's not just an African problem: 12 species of mammal have disappeared from Vietnam's forests since 1975.

Yet banning bushmeat, as the flesh of wild animals is called, would deny forest dwellers an important source of protein. "You can't say 'don't hunt', because a lot of rural people depend on it," says Elizabeth Bennett, director of the Hunting and Wildlife Trade Program at the New York-based Wildlife Conservation Society (WCS). So rather than eliminating the bushmeat trade, conservationists and ecologists want to manage it.

The theoretical models used to analyse the impacts of hunting must be quick and easy to use on the ground — where most forest animals lead secretive lives hidden from ecologists — but also have to cover the gamut of species and environments. The results are rough, and ecologists differ over what are the best methods. But the approach is not hopeless: some damaging hunts have been curbed, whereas others seem to be sustainable, allowing conservation efforts to be focused on where they're needed most. And ultimately, conservationists have a duty to try and understand hunting, argues Richard Bodmer of the University of Kent in Canterbury, UK. "If you can't provide



Hung out to dry: conservationists fear that many forest mammal species will be wiped out by hunters.

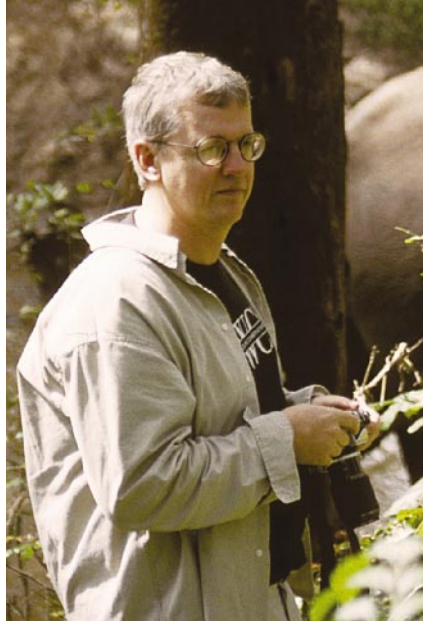
communities with information on how many animals they can hunt, you've failed."

John Robinson and Kent Redford, ecologists with the WCS, devised the most commonly used method for assessing the impact of bushmeat hunting¹. Their model requires only an estimate of the population of a particular species in the region under study. This is usually inferred from counts of dung, nests, or sightings of animals made at a set of fixed points. Other parameters, such as a species' reproductive rate and lifespan, can be obtained from previous studies or calculated from body size. From these figures, the model gives the theoretical maximum sustainable harvest. "It shows you the best of all possible worlds," says Robinson. This is then compared to the actual harvest — which can be measured by interviewing hunters — to see whether a hunt is sustainable or not.

Robinson and Redford's analysis, which was published in 1991 and is used in Asia, Africa and Latin America, paints a stark picture. Tropical-forest mammals, say the duo,

live at such low densities and breed so slowly that their meat can support only about one human hunter per square kilometre. When applied to the Malaysian province of Sarawak, the model showed that any commercial trade in wild meat — pigs, monkeys, deer and rodents, for example — was unsupportable. The WCS results persuaded the Malaysian government to ban all except subsistence hunting in Sarawak. Heather Eves, director of the Bushmeat Crisis Task Force, a consortium of conservation organizations and scientists based in Silver Spring, Maryland, believes that similar action is needed worldwide. "The tools we have today show that, across the board, commercial hunting of wildlife is unsustainable," she says.

But other models indicate that some animals can be, and are being, hunted at safe levels. Bodmer has developed the 'unified harvest model', a more detailed approach that builds on Robinson and Redford's work and combines several methods for analysing animal ecology and hunting^{2,3}. "Using more than



Kent Redford (above, left) and John Robinson want to assess the sustainability of current hunting.

one model will increase your confidence,” says Bodmer — if the models agree, there’s a better chance that researchers are on the right track.

The unified harvest model generates two numbers: the level at which a hunted population must be maintained for it not to decline, and the proportion of animals born each year that can be safely taken. Baseline population data come from areas where there is little or no hunting. Comparison with hunted areas shows how much the population has been depleted, and the number of pregnant animals killed gives a guide to the species’ birth rate. As a ballpark figure, says Bodmer, the harvest of large, slow-breeding species should comprise only 20% of the newborns. For fast breeders, a harvest of 40% should be sustainable.

Working with communities in the Peruvian Amazon, Bodmer has calculated that the region’s wild pigs, deer and large rodents seem to be standing up to the current level of hunting. Last July, at the annual meeting of the Society for Conservation Biology, held at the University of Kent, he showed that the more of a species that are killed, the more offspring the survivors produce, probably as a result of reduced competition for food. But forest

inhabitants that have not evolved to cope with predators, such as the large, slow-breeding lowland tapir (*Tapirus terrestris*), are less flexible, and Bodmer advises against hunting them. He is optimistic that his message is being heard, saying that forest dwellers are keen to manage their hunting.

The methods used both by Robinson and Redford and by Bodmer are intended to assess existing hunts, rather than to guide future hunting — although Robinson says that his model has sometimes been used inappropriately to set harvest levels. The analyses can only show that hunts that exceed a theoretical maximum are definitely unsustainable; harvests that fall below this level might be damaging or they might not.

But some ecologists think that the methods cannot achieve even this modest goal. Eleanor Milner-Gulland of Imperial College, London, points out that errors of 30% in population estimates are common, and that methods that use these estimates to set hunting thresholds will mislead. “As soon as you subject the models to anything like realistic uncertainty they fall down,” she says. “Thresholds for sustainable hunts go much lower if you incorporate

uncertainty and bias.”

Milner-Gulland wants hunting studies to take explicit account of our ignorance, and is developing alternative models that use Bayesian statistics. Commonly used in fisheries science, these give results as a distribution of probabilities, rather than as a single number⁴. But balancing accuracy with usability can be tricky. Some researchers feel that it is important to create models that non-scientists from development agencies can use (see “A market solution?”, below). Others wonder whether models such as Milner-Gulland’s will be too difficult to apply. “Our model was very quick and dirty,” admits Robinson, “but going to the other extreme isn’t going to be very useful.”

As well as being vulnerable to uncertainty, simple models can fail to account for biological reality. Last August, Philip Stephens of the University of Wyoming in Laramie and his colleagues described how ten different models fared at predicting the population dynamics and sustainable harvest from an intensively studied colony of marmots (*Marmota marmota*)⁵ in Germany. Robinson and Redford’s model overestimated the harvest, largely because it ignores social behaviour. Animals that live in groups, including marmots and many primates, can become extinct if their group size falls below a critical level. “You get a very sharp change between safe and dangerous levels of exploitation,” says Stephens.

Given these problems, can ecological models ever hope to guide policy? Conservationists are aware of the deficiencies, but insist that models have a valuable role to play. As forests shrink and human populations grow, the best long-term hope for wildlife is to provide local people with alternative food sources. But in the meantime, understanding and regulating hunting is essential, as the projects in Malaysia and the Amazon have shown.

Science can’t solve the problem, but those involved are adamant that it is part of the solution. And it needs to be implemented quickly. “We know very little about the ecology of most hunted species, but we know enough to be able to get some management on the ground,” says Bennett. “The problem is so acute that if we wait it’ll be too late.” ■

John Whitfield works in Nature’s news syndication team.

1. Redford, K. H. & Robinson, J. G. in *Neotropical Wildlife Use and Conservation* (eds Redford, K. H. & Robinson, J. G.) 415–429 (Univ. Chicago Press, Chicago, 1991).
2. Robinson, J. G. & Bodmer, R. E. *J. Wildl. Mgmt* **63**, 1–13 (1999).
3. Bodmer, R. E. & Robinson, J. G. in *People in Nature: Wildlife Conservation in the Neotropics* (eds Silviu, K., Bodmer, R. E. & Fragoso, J.) (Columbia Univ. Press, New York, in the press).
4. Milner-Gulland, E. J. & Resit Akçakaya, H. *Trends Ecol. Evol.* **16**, 686–692 (2001).
5. Stephens, P. A., Frey-Roos, F., Arnold, W. & Sutherland, W. J. *J. Anim. Ecol.* **39**, 629–642 (2002).
6. Clayton, L., Keeling, M. & Milner-Gulland, E. J. *Ecol. Appl.* **7**, 642–652 (1997).

Bushmeat Crisis Task Force

■ www.bushmeat.org

A market solution?

Persuading governments to implement conservation plans currently takes years of field-work, so Guy Cowlshaw of the Institute of Zoology in London is trying to speed things up. His team is studying the bushmeat chain from hunter to market. Once this is understood, he says, it may be possible to work out what’s going on in the forest by perusing a market’s stalls.

Cowlshaw’s team is looking at Ghanaian hunters’

taste in prey, and the point at which they switch to smaller species as big game becomes rare. “If you see a market with lots of small animals, it’s almost certainly been over-exploited,” he says. The aim is to develop bioeconomic models that can measure the effects of hunting from information on price and availability. “We’re trying to come up with simple rules of thumb that people in development agencies

can use,” says Cowlshaw.

The model has its critics, however. Eleanor Milner-Gulland, an ecologist at Imperial College, London, has found that, for wild-pig hunting in Indonesia, the market stayed the same while the population collapsed, as hunters simply went further in pursuit of game⁶. Animals eaten in the forest, and those discarded *en route* to market, can account for more than half of those killed, she adds.

A window on the inner Earth

Seismic studies of the deep roots of Hawaii's volcanoes may help to reveal mysterious circulation processes in the Earth's mantle — shedding light on our planet's history and dynamics. Rex Dalton peers beneath the surface.

For most of us, the Hawaiian Islands are a holiday hotspot, a haven for sun and surf. But for Earth scientists, the archipelago's real appeal lies in the geophysical hotspot that lies beneath the islands and is responsible for their formation.

Next autumn, a multi-institutional team will encircle Hawaii with an array of seabed seismometers to study the hotspot's roots deep within the Earth's mantle. The \$2.5-million project, backed by the US National Science Foundation (NSF), is called PLUME, for Plume-Lithosphere Undersea Melt Experiment. By recording seismic waves from distant natural earthquakes, researchers hope to chart the hotspot's plumbing — and in doing so address fundamental questions about the processes that have shaped, and continue to shape, our planet.

Much of the Pacific Ocean's sea bed is a single plate that, over the past 70 million years, has drifted slowly northwest relative to the layers beneath it. The hotspot is the result of a rising plume of hot buoyant rock, which deforms the overlying tectonic plate and spreads out beneath it. As the plate has slowly moved over the plume, this has created a huge plateau, or 'swell', some 2,000 kilometres long and half as wide (see Diagram, opposite).

The solid rock rising up through the plume is released from the enormous pressure of the Earth's interior and starts to melt. This in turn melts some of the overlying plate, or lithosphere, and the resulting mixture can then seep through cracks in the plate, giving rise to volcanoes. The plate's movement has produced a string of volcanic islands stretching away from the hotspot. Today, the plume is centred to the southeast of Hawaii's Big Island, near the Loihi seamount, an undersea active volcano.

The volcanic processes that gave rise to the Hawaiian Islands are relatively well understood. But the plume that feeds the hotspot remains shrouded in mystery. Researchers want to know the plume's width, the temperature of the rock rising up through it, and more about its geochemical composition. Most fundamentally, they want to know how deep in the Earth's mantle are the plume's origins. "The mantle has been like a mysterious black box," says



Chain reaction: the Hawaiian Islands stretch away from the geophysical hotspot that formed them. Currently, the hotspot is centred just southeast of the Big Island (bottom right).

Naomi Oreskes, a geologist and historian of Earth science at the University of California, San Diego. "Finally, technology is allowing the opening of that box."

Core issues

Long-lived hotspots such as that beneath Hawaii — others underpin Iceland, the Galapagos Islands and the Cape Verde Islands, west of Africa — are believed to depend on stable plumes that arise from convection currents originating deep within the mantle. But it remains unclear from what depth the plumes arise. They may have their roots in the upper mantle, less than 660 km below the surface. But if their origins go deeper than this, as many geophysicists believe, then they probably come from about 2,900 km, at the boundary between the lower mantle and the seething cauldron of iron-rich material that forms the Earth's core.

For geophysicists, the plume's depth is an important issue. Leading models of the circulation of material in the Earth's mantle — which drives the movement of the lithosphere's tectonic plates — result in plumes

transferring material across the boundary between the upper and lower mantle, at a depth of 660 km. But at this point, hot rising minerals undergo a phase transition that results in a decrease in temperature and therefore a decrease in buoyancy. Some geophysicists speculate that this may be sufficient to prevent hot rock, rising from the core-mantle boundary, from punching straight through to the upper mantle — so they argue that hotspots observed at the Earth's surface must have a shallower source.

If plumes do not traverse the transition zone at 660 km, it would also hinder researchers interested in the lower mantle's composition — which, in turn, informs our understanding of how the planet formed and has evolved over the eons. Geochemists have made inferences about the Earth's deep interior by analysing material that has reached the surface in volcanic eruptions over long-lived hotspots. If none of this material comes from near the core-mantle boundary, they will have to revise their ideas.

PLUME should remove these unknowns by providing the best picture yet of the 660-km transition beneath a hotspot — and so

NASA/SPL

it will be watched keenly by Earth scientists worldwide. The project is significant also because it represents the most ambitious deployment of undersea seismometers yet attempted. "This experiment will be used as a way to rethink what marine seismology is all about," asserts Gabi Laske, a seismologist at the Scripps Institution of Oceanography in La Jolla, California, and a principal investigator for PLUME.

Action stations

In September, the research vessel *Roger Revelle* will deploy the first PLUME instruments, placing 35 seismometers on the ocean floor at depths of 4–6 km. For the first 15 months, the seismometers will be placed some 75 km apart, surrounding the Big Island. During the second 15-month deployment, the array will be enlarged around the entire Hawaiian Island chain, with the instruments set some 200 km apart. A further 10 seismometers will be placed on the islands themselves throughout the project.

The seismometers will record the precise timing and amplitudes of seismic waves. For instance, if an earthquake occurs in Tonga — some 5,000 km south of Hawaii — PLUME's instruments will pick up the seismic waves as they ripple through the Earth. The timing of their arrival at Hawaii is crucial: by knowing exactly when the wave arrives at each seismometer, researchers can determine the speed with which the waves propagated through the region below the seismic array. With this information, the team will infer what the waves passed through — for example, seismic waves move more slowly through

Enthusiasts for marine seismology believe that it is poised to mine a rich vein of discoveries.

hot rock than through cooler material.

By combining information from many earthquakes, PLUME investigators will draw up a three-dimensional picture — in much the same way that X-ray images are combined to form the familiar three-dimensional computerized tomography scans used in diagnostic medicine.

The PLUME team members believe that clustering their seismometers as proposed will produce a detailed picture of the plume down to almost 700 km. If it extends to these depths without widening substantially, the experiment would provide some of the strongest evidence yet that it emanates from within the lower mantle, most likely the core–mantle boundary.

Some researchers believe that it is possible to peer deeper into the Earth's interior than the PLUME team is planning, by using more widely spaced instruments and applying various computational tricks to tease more information from the seismic data. In 1999, for instance, Harmen Bijwaard and Wim Spakman of Utrecht University in the Netherlands claimed to have imaged the plume beneath Iceland right down to the core–mantle boundary¹ — albeit at much

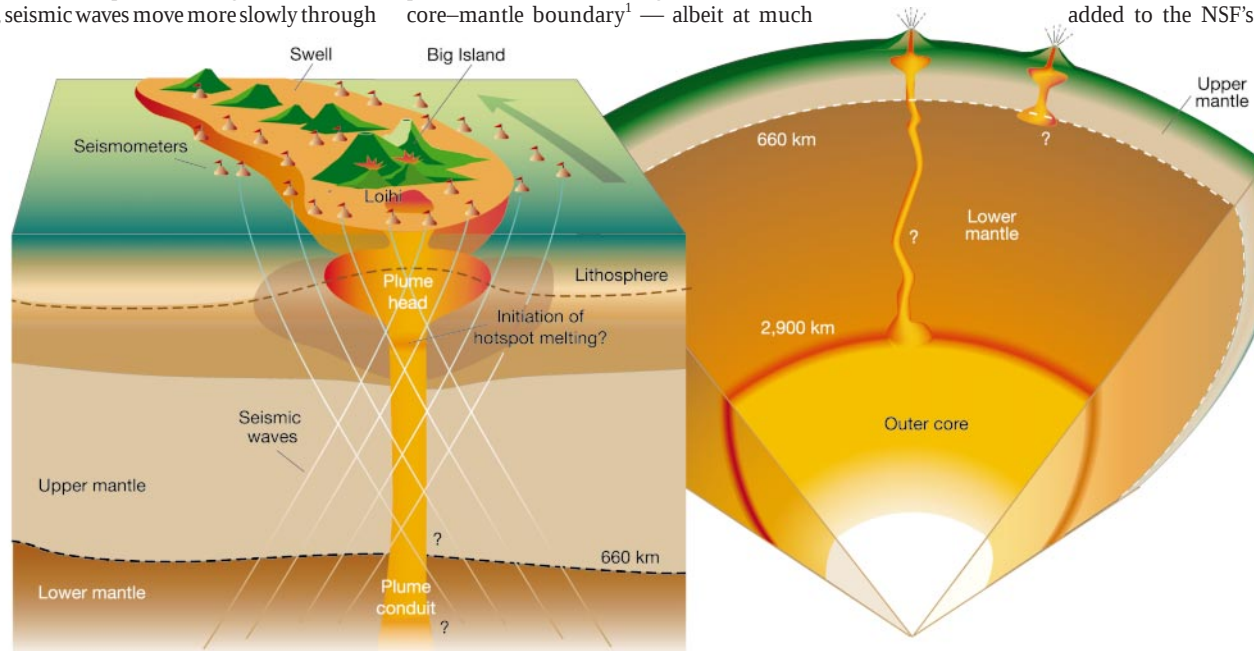
lower resolution than the PLUME team hopes to achieve. But although many experts agree that Bijwaard and Spakman's results support the idea that there is a deep plume beneath Iceland, the method pushes the data to their limits, and the case is far from closed.

PLUME will help to clarify the picture — and it might be possible to obtain deeper images by placing PLUME's seismometers farther away from the Hawaiian Islands. "But we then would lose the ability to see the small details that allow us to determine where the plume is," says Laske, who believes that the configuration planned for PLUME offers the optimal trade-off — given the current number of seismometers available — between image depth and resolution.

Dynamic drive

As the PLUME team gears up for its deployment, a broader group of researchers is coming together to follow the trail that Laske and her colleagues are blazing. Last September, they held a workshop in Snowbird, Utah, to prioritize projects under the NSF's Oceanic Mantle Dynamics Science Plan. Conceived in 2000, this is a blueprint for interdisciplinary studies of the deep geological structures under the world's oceans. Devising a method to produce high-resolution seismic images of plumes in the lower mantle, all the way down to the core–mantle boundary, is one of its goals.

The plan has yet to be funded, but its prospects will receive a significant boost should PLUME produce the impressive results that its proponents anticipate. Once PLUME is over, its instruments will be added to the NSF's



Depth charge: a conduit of hot buoyant rock rises up from beneath the Hawaiian Islands to form an undersea plateau, or swell. As the lithosphere has moved across this hotspot, molten rock has occasionally broken through the surface to form volcanic islands. Researchers now aim to investigate whether the source of the plume lies in the upper mantle or whether it cuts right through the mantle from the boundary with the core (right).

pool of equipment for undersea seismology — the availability of which is the main constraint on the field's development. "With new technologies, we can do experiments that were not possible before," says John Orcutt, a geophysicist at Scripps and a principal investigator for PLUME. "But they tie up instrumentation for a long period of time."

Sink and swim

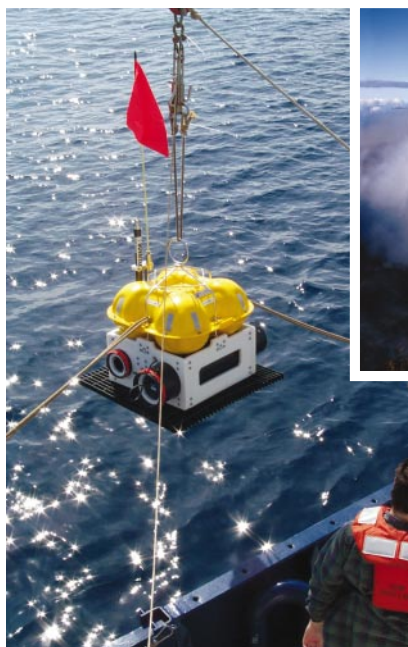
PLUME has been made possible by durable seismometers that can be dropped to the ocean floor, record data from earthquakes over many months, and then float to the surface to be recovered for redeployment. Those to be used for PLUME were developed by Orcutt from inexpensive prototypes dubbed 'L-Cheapos'. The current versions are more sophisticated, and cost \$25,000 each. For a while, researchers knew them as 'L-Expensivos'. But the Scripps team eventually decided that this moniker was too frivolous, and renamed them LC2000s.

Each LC2000 is mounted on a concrete base that sinks it to the ocean floor. The base is all that is left behind after an acoustic signal prompts the equipment to ascend for retrieval. The current devices' improved robustness, along with technological advances in data storage and battery power, mean that they can be used for successive extended deployments.

The LC2000s can also record seismic waves with longer periods than was possible with the seismometers borne by the earlier L-Cheapos, which were blind to waves with periods longer than 45 seconds. Previously, researchers used differential pressure gauges to detect long-period waves, but these instruments don't record the direction from which the waves have come. The LC2000s should overcome this obstacle for waves with periods as long as 80 seconds. "This is the first time these will be deployed," says Laske. "They are very promising."

In 1998, some members of the PLUME team conducted a proof-of-principle experiment called Seismic Wave Exploration of the Lower Lithosphere, or SWELL. Led by Laske, Orcutt and Jason Phipps Morgan, now at GEOMAR, the Research Center for Marine Geosciences in Kiel, Germany, SWELL involved eight L-Cheapo seismometers placed 200 km apart in a hexagonal array around part of the Hawaiian swell. The data obtained revealed the western edge of the plume beneath the swell, and showed that it comes from a depth of at least 200 km (ref. 2). "SWELL showed it was technically feasible to do this type of experiment under the ocean," says Laske.

But PLUME is not just about seismology. As Laske and her colleagues chart the conduit of rising hot rock under Hawaii, geochemists on the team will be considering the plume's relationship with the atmosphere and oceans.



Taking on the mantle: by deploying 35 LC2000 seismometers (left) around Hawaii, the PLUME team hopes to gain insight into the processes powering volcanoes such as Kilauea (above).

For instance, studies of volcanic gas trapped in bubbles in rock at volcanoes on the Hawaiian islands, and dissolved in water from hydrothermal vents on the Loihi seamount, have shown that the ratio of the isotope helium-3 to helium-4 is much higher — 30 times at Mauna Loa on the Big Island, and 20 times at Loihi — than in the atmosphere³. It is also higher than in the gas dissolved in water from hydrothermal vents at the mid-ocean ridge, which is thought to come from the uppermost region of the mantle.

It's a gas

Geochemists theorize that these elevated ratios show that the plume is bringing up rock from deep within the mantle that previously could not emit its helium-3. But proof awaits confirmation of the depths to which the plume extends, and more detailed geochemical analyses. "PLUME will give us clues to the way the Earth emits gas," says David Hilton, a geochemist at Scripps. And because the evolution of our atmosphere and oceans has depended heavily on gases emitted from the solid Earth, such clues are fundamental to understanding our planet's history.

Erik Hauri of the Carnegie Institution of Washington, who is leading the geochemical component of PLUME, hopes to deepen our knowledge by examining three major volcanoes on the Big Island of Hawaii: Mauna Loa, Kilauea and Mauna Kea. They are all within 60 km of one other, yet the composition of their lava is subtly different⁴. "I want to try to define what parts of the plume are giving rise to what volcanoes," says Hauri. This, in turn, should help seismologists to understand the labyrinth of conduits that may rise up to the surface from the plume's head.

While PLUME's principal investigators

prepare for action, supporters of the NSF's Ocean Mantle Dynamics Science Plan must also make themselves busy. For if they are to gain the maximum benefit from their proposed experiments, they need to get funding in place without delay. They are looking for some \$20 million to build 350 ocean-floor seismometers, plus running costs of about \$10 million a year.

At the Snowbird workshop, participants decided that a top priority was to complete a seismic analysis of the United States' continental shelves out to at least 600 km along the west, east and gulf coasts. And they agreed that this will be most informative if conducted in parallel to a project called the USArray, in which portable seismometers will be deployed across the United States over the next decade, to image structures deep beneath the continent⁵. "You want to record the same earthquakes," Laske explains.

The USArray is scheduled to begin deployment in Southern California in 2005. The ocean mantle team would like to have its seismometers offshore by 2006. "It's going to be difficult," says Robert Detrick, a seismologist at the Woods Hole Oceanographic Institution in Massachusetts. "But if we really want to take advantage of the USArray timing, we should do it then."

Enthusiasts for marine seismology believe that their nascent field is poised to mine a rich vein of discoveries. "The plan is to look much deeper into the Earth," says Orcutt. "A lot of us feel the need to get on with it." Members of the PLUME team, meanwhile, hope that their project will provide an exciting preview of what is to come. "We've wanted to do an experiment like this for 20 years," says Hauri. "It'll be pretty cool."

Rex Dalton is Nature's US West Coast correspondent.

1. Bijwaard, H. & Spakman, W. *Earth Planet. Sci. Lett.* **166**, 121–126 (1999).
2. Laske, G., Phipps Morgan, J. & Orcutt, J. A. *Geophys. Res. Lett.* **26**, 3397–3400 (1999).
3. Kurz, M. D., Jenkins, W. J., Hart, S. R. & Clague, D. *Earth Planet. Sci. Lett.* **66**, 388–406 (1983).
4. Hauri, E. H. *Nature* **382**, 415–419 (1996).
5. Dalton, R. *Nature* **405**, 390–392 (2000).

Scientific research and the human condition

Reviewers are only human, even though they wield enormous power for a moment.

Sir — Today I am reviewing an article for a scientific journal. I will generate criticism and express my opinions, and the editor will probably accept them without hesitation, as I am an expert reviewer, if only a god for an instant. But I am only a man, full of inconsistencies and capable of misunderstanding, misrepresenting and making fatal errors.

These facts are often overlooked in the review process. The monstrous machinery we have constructed around scientific research, of which peer review is just one example, does not always augment the quality of research nor make things easier for the average busy researcher who has to fight for funding and prestige in the scientific community (see “Publish and be damned...”, *Nature* **419**, 772–776; 2002). After all, fundamental science was done in

ancient times, long before the advent of peer review and journal impact factors.

I wonder to what extent my feelings are common to other members of the scientific community when I receive manuscripts or grants for review, and hear a distressed inner voice asking who am I to tell these investigators that what they have done, or want to do, is of no interest to most of the community, and that their work should not be published, or their grant not funded? Whoever is really into research has a modest ego, as we understand our transient existence in a mostly indifferent Universe.

The act of judging — specifically when personal views have to be exposed, such as the ‘relevance’ I attribute to a particular study for publication in a specific journal — becomes increasingly difficult as I reflect on these issues. The consequences

of our extremely involved scientific machinery, and whatever standards we set in our individual or communal research endeavours, are limited by our human condition. We are uncovering fraudulent data, fictional experiments proposed in grant applications that are never performed, rivalry, mistrust and other mischievous behaviour. One sad consequence is the loss of some creative individuals who decide to quit research.

How can we all, as individuals, contribute to a community effort in which everyone — students, technicians and professors alike — can enjoy the pleasures of unravelling the mysteries of nature?

Jose Luis Perez Velazquez

Hospital for Sick Children, Division of Neurology,
555 University Avenue, Toronto,
Ontario M5G 1X8, Canada

Malaria — there could be a third way

Sir — The excitement of the simultaneous publication of the malaria parasite *Plasmodium falciparum*¹ and mosquito vector *Anopheles gambiae*² genomes is contagious. Although it is to be hoped that new drugs for malaria are now around the corner, there are some chilling phrases in *Nature*’s bumper issue on *Plasmodium* genomics³. We read that “more people are infected with malaria in Africa today than at any time in history”, and that it will require “perhaps two decades to convert genomic information into effective vaccines”. A more optimistic view is that all we need to achieve a vaccine is better delivery of known antigens, for example by heterologous priming and boosting.

A stated goal for an effective malaria vaccine is to “induce protective immune responses equivalent to, or better than, those provided by naturally acquired immunity or immunisation with attenuated sporozoites”. Why, then, is there no attenuated sporozoite vaccine? Part of the answer is that sporozoites cannot easily be grown *in vitro*. Proteomic analysis, with small quantities of sporozoites from infected mosquitoes, has neatly resolved the problem of analysing sporozoite gene expression. Astonishingly, this has shown that sporozoites express many genes of the *var* and *rif* families, which are largely responsible for antigenic variation. This is encouraging for vaccine development, and poses several obvious questions. Are the antigen genes expressed in sporozoites

representative of *var* gene diversity or are they at specialized genomic sites? Do the sporozoite antigens change with sequential cycles of a single strain? How divergent are the sporozoite repertoires of different strains?

The *var* and *rif* antigenic diversity of sporozoites could stimulate new investment in seeking cultivation technology for the entire *Plasmodium* life cycle. The *P. falciparum* and *A. gambiae* genomes are littered with clues as to what the parasite needs to thrive. There have been some notable preliminary successes with *in vitro* systems. A relatively simple culture system³ has been described for growing oocysts and sporozoites of *P. falciparum*, but does not seem to have been exploited. More recently, a similar system has been achieved for *P. berghei*⁴.

It would be foolish to suggest that all that is needed to obtain a malaria vaccine is to grow sporozoites in bulk. The protection attained with X-irradiated sporozoites, or X-irradiated infected mosquitoes, was a benchmark, but resistance faded and challenge broke through in some volunteers⁵. *Var* gene diversity is generated by recombination, and the repertoire of the blood-form stages appears to be immense⁶.

Targeted regional vaccines are unlikely to work because antigenic types do not cluster geographically. There is no guarantee that sporozoite pools can be generated with adequate antigenic repertoires, although immune pressure *in vitro* might be used to improve the yield of *var* gene recombinants. Nevertheless, a determined sporozoite cultivation consortium initiative would cost little and should pay rich dividends. New

means of sporozoite attenuation and delivery could be explored.

Michael A. Miles

Department of Infectious and Tropical Diseases,
London School of Hygiene and Tropical Medicine,
Keppel Street, London WC1E 7HT, UK

1. Gardner, M. J. *et al.* *Nature* **419**, 498–511 (2002).
2. Holt, R. A. *et al.* *Science* **298**, 129–149 (2002).
3. Warburg, A. & Schneider, I. *Exp. Parasitol.* **76**, 121–126 (1993).
4. Al-Olayan, E. M. *et al.* *Science* **295**, 677–679 (2002).
5. Richie, T. L. & Saul, A. *Nature* **415**, 694–701 (2002).
6. Fowler, E. V. *et al.* *Mol. Biochem. Parasitol.* **120**, 117–126 (2002).

Academics are teachers and colleagues too

Sir — The proposal in Correspondence by E. Fernández-Juricic *et al.* (*Nature* **420**, 16; 2002) to apply impartial peer review to academic hiring ignores several important factors. One is the central role of academic institutions — to educate.

At most academic institutions, faculty members teach as well as conduct research. Merely assessing research qualifications is no indication of the ability to communicate complex ideas or to empathize with students. Personality also looms large, particularly in small departments — the highly qualified eccentric who cannot interact with people may be a difficult and ineffective colleague.

In their zeal to provide dispassionate assessment, Fernández-Juricic *et al.* seem to have forgotten that science is a human activity. My hope is that their proposal was made tongue-in-cheek.

Karen F. Greif

Department of Biology, Bryn Mawr College,
Bryn Mawr, Pennsylvania 19010, USA

1903 and all that

Honours go to the first powered flight and the DNA double helix.

J. L. Heilbron and W. F. Bynum

"If DNA was so important, why weren't you working on it?" Mrs Linus Pauling asked her husband sometime after the revolution. In fact, many people in the early 1950s worked on the structure of DNA, increasingly identified as the material basis of heredity. The allocation of the spoils is still a parlour game, as a close guess at the structure was made by Pauling, and a key piece of information, an X-ray analysis of the DNA molecule by Maurice Wilkins and Rosalind Franklin, contributed immediately and essentially to the discovery. But there is no doubt that the structure of life was first revealed in this journal, in the 25 April issue of 1953, by a US postdoc at Cambridge, James Watson, and his British colleague, Francis Crick. Their helical model of DNA is the most famous scientific icon of modern times, and their short communication of 1953 is our choice for anniversary of the year.

In the 15 years that we have been communicating our anniversarial researches to readers of *Nature*, our conventions have remained constant. As usual, ¢ signifies a century, \$ a Nobel prize; thus Watson (*b.* 1928, 0.75 ¢), Crick and Wilkins, \$1962. (Franklin could not have figured as she had died before Stockholm honoured the discovery.) Maintaining tradition, we move back by centuries and half-centuries, recognizing at the end the importance of more modern scientific discoveries by covering the twentieth century in 25-year intervals.

1903 (1.0 ¢)

You will have to wait until the end of the year to celebrate the most famous of this year's centenaries. On 17 December, Orville Wright made an 853-foot flight taking just 59 seconds at Kitty Hawk, South Carolina. By 1908, Orville's brother, partner and co-inventor Wilbur was astonishing European aviators by flying around France for 2 hours and 20 minutes non-stop. The same year that these former bicycle-makers demonstrated the first-ever powered, manned and controlled heavier-than-air-flight, a half-deaf Russian school teacher, Konstantin Eduardovitch Tsiolkovsky, was designing heavier-than-vacuum travel. He foresaw reactive propulsion, space suits, satellites, space stations and multi-stage rockets.

Only the fit and stout of heart can be astronauts. Providence therefore arranged that as Tsiolkovsky was writing about space travel, Willem Einthoven (\$1924), professor of medicine in Leiden, The Netherlands, per-



James Watson (left) and Francis Crick's discovery in 1953 of DNA's structure changed the face of science.

fected his string galvanometer for mapping cardiac activity. In 1903 he set out precise protocols for making electrocardiograms. Subsequently, by means of a cable 1.5 kilometres long, he hitched up heart patients in the academic hospital to the machine in his laboratory, and gradually worked out the relationship between the electrocardiogram and particular damage or malfunctions of the heart. In the 1920s, in collaboration with his son, Einthoven adapted his string galvanometer to receive radio messages from the Dutch colony of Java, where he had been born. He could thus receive signals from the great continuous-wave arc generator erected there in 1921 on principles established by the Danes Valdemar Poulsen and Peder Oluf Pedersen 100 years ago.

Among those who died in 1903 were Richard Jordan Gatling, American inventor of the Gatling gun; Josiah Willard Gibbs of Yale, deep-thinking theoretical thermodynamicist and statistical mechanic; the philosopher and sociologist Herbert Spencer; and George Gabriel Stokes, the Cambridge mathematical physicist. The German mathematician Gottlob Frege survived 1903, but his life's work, *Grundgesetze der Arithmetik* (two volumes, 1893–1903) did not. It was overturned just before its completion by Bertrand Russell,

who pointed out a logical contradiction in its foundation. Russell himself might deserve a centennial party, not for undoing Frege but for completion of the first volume of his classic *The Principles of Mathematics*.

For our money, the prize for 1903 should go to King C. Gillette, who brought his safety razor to market after several years of research and development. In 1903 he sold just 51 razors and 168 blades. The next year he sold 123,648 blades, an increase of exactly 736%. Almost as meritorious was the decaffeination process conceived by the German coffee importer Ludwig Roselius after inadvertently soaking coffee beans in salt water. He marketed the result as SANKA (for *sans-caféine*).

1853 (1.5 ¢)

Along with the Wright brothers' achievement of 1903, the thrifty celebrant will want to note George Cayley's monoplane glider, launched from a balloon in Yorkshire in 1853. Cayley (*b.* 1773, 2.3 ¢) had been working on powered flight off and on since the 1790s, but the task of running his estates kept him grounded until he was too old to fly. His reluctant coachman was the pilot in the first attempt, which glided 500 yards.

Other persuasive parallels between 1853 and 1903 present themselves. Poulsen is

A. BARRINGTON BROWN/SPL

paralleled by Latimer Clark, who built a pneumatic telegraph 220 metres long between the London Stock Exchange and the International Telegraph Company; Frege by the unfortunate William Shanks, who erred in the 528th and hence all subsequent places in his computation of π to 707 decimal places; Gillette by Amédée Bonnet of Lyon, who improved the amenities of life by establishing a rational regime of massage; and Roselius by George Crum, of Saratoga Springs, New York, who, in trying to reproduce Parisian *frites*, invented the potato crisp.

Léon Foucault demonstrated with great ingenuity that light travels more slowly in water than in air, which no one doubted; whereas John Couch Adams showed with equal dexterity what no one had dared to think — that Laplace's theory of the Moon's motion was incomplete. Charles Frédéric Gerhardt published the first volume of his *Traité de Chimie Organique*, which transmitted the mature form of his theory of types, a classificatory scheme based on the form rather than the content of chemical formulae.

Among those known to science who died in 1853 were Dominique-François Arago, expert in optics, astronomy and electromagnetism, and also for many years a deputy to the French Parlement and a standing secretary of the Académie des Sciences in Paris; William Beaumont, American surgeon, whose observations through a patient's gastric fistula resulted in his celebrated *Experiments and Observations on the Gastric Juice and the Physiology of Digestion* (1833, 1.7 €); Christian Leopold von Buch, staunch defender of neptunism, which explains all major geological features as originating from the actions of water; and Auguste Laurent, Gerhardt's comrade in the development of type theory. These deaths were partially offset by the birth in Wisconsin of the American-British pharmaceutical entrepreneur Henry Solomon Wellcome, whose trust is a major benefactor of contemporary biomedical research.

1803 (2.0 €)

No fewer than five new elements turned up in 1803. Jöns Jacob Berzelius and his colleague Wilhelm Hisinger found cerium as, independently, did Martin Heinrich Klaproth (b. 1743, 2.6 €). Smithson Tenant and William Hyde Wollaston, his partner in platinum refining, each discovered two elements: iridium and osmium, and palladium and rhodium, respectively. The names of these elements, all derived from Greek, either indicated physical properties or referred to contemporary astronomical discoveries. Iridium (from *iris*, a rainbow, after the range of colours of its solutions), osmium (from *osmis*, odour, after the smell of its tetrachloride), and rhodium (from *rhodos*, reddish, after the colour of its ammonium chloride) belong to the first group. Cerium and palladium take their names from the asteroids Ceres and Pallas,



Lift off: Orville Wright, seen lying on the plane's lower wing, undertook the first powered flight in 1903.

which were discovered in 1801 and 1802 (and named for the Greek goddesses Ceres and Pallas Athene). This second group of elements named after heavenly bodies was not enlarged until the Berkeley cyclotron produced the trans-uranic elements neptunium and plutonium in 1940. The discovery of the asteroids had another immediate consequence in helping Jean Baptiste Biot to argue persuasively, in 1803, that meteorites come from the apparently plentiful stock of rocks circulating through the Solar System.

As the number of elements, in Lavoisier's sense of bodies not (yet) decomposable chemically, increased, so did the desire to classify and rationalize them. By chance or design, a paper on the absorption of gases by liquids, which appeared as the five metals of 1803 were being identified, contained the seeds of the required system. It was John Dalton's first "enquiry into the relative weights of the ultimate particles of bodies, [which] is a subject, as far as I know, entirely new". This first sketch of Dalton's atomic theory discussed only particles of permanent gases. Gradually the metals fell into the scheme and those of 1803 found comfortable accommodation in Mendeleev's first periodic tables.

Two centuries ago Claude-Louis Berthollet and Gaspard Monge saw their famous

books, *Essai de Statique Chimique* and *Géométrie de Position*, respectively, through the press. We hope that while doing so they spared some time to watch Robert Fulton's steamship make its first voyage on the Seine.

Chemistry received a further kick-start through the birth of Justus Liebig. Yet another first appearance was that of Johannes Christian Doppler, in Vienna. Thirty-two years later, when on the verge of emigrating to the United States, he obtained a professorship of mathematics in Prague. There, in 1842, he announced the Doppler effect, with respect not to sound but to starlight. Its application to acoustics was confirmed in Utrecht in 1845 by Christoph Hendrik Diederik Buys Ballot, who used a railway car full of trumpeters to achieve the expected effect.

1753 (2.5 €)

Could all seven of the bridges linking Königsberg's two islands to the rest of the Finnish city be crossed in a single journey without going over any of them twice? "No," said Leonhard Euler, 250 years ago, and proved it. The same year, for longer trips, James Lind advised, in his *Treatise on Scurvy*, that citrus juice was as useful as mathematics on long voyages.

Somehow, while sending early volumes



Classified information: John Dalton laid the foundations of the periodic table of elements in 1803.

of his *Encyclopédie* to the printer, Denis Diderot managed to write his *Pensées sur l'interprétation de la Nature*, which opens with the news that mathematics as a tool for exploring nature had come to a dead end. That was true, to be sure, of the mathematical philosopher George Berkeley, who had demonstrated the logical fallacies in Newton's calculus, and the mathematically inclined professor of physics at the Petersburg Academy of Sciences, Georg Wilhelm Richman. Richman died the most celebrated death of any enlightened philosopher. A stroke of lightning ended his short career as he attempted to demonstrate its identity with electricity and before he could read the explanations that might have saved his life in Giambattista Beccaria's franklinist *Dell'Elettricismo Artificiale e Naturale* (1753).

1703 (3.0 ¢)

The big news in science three centuries ago was the election of Isaac Newton as president of the Royal Society of London, a position he held until his death in 1727. Under Newton the society returned to its early practice of showing and witnessing experiments at most of its meetings. The work was done by Francis Hauksbee, who in the year of Newton's accession began a series of experiments

on glows in evacuated vessels rubbed against the hand. While thus engaged, Hauksbee invented the prototype of the electrostatic generator, which became the emblem of enlightened experimental philosophy. Although experiments constituted most of the business and purpose of the society, the fellows, their president included, did not disdain the telling of tall tales. Newton once capped a story of a lady who had grown teeth at the age of 64 with the announcement that he knew a crone who had done as much at 90.

One reason that Newton became available for the presidency in 1703 was the death in March that year of his rival Robert Hooke, who had challenged his ideas about light and colour, and his priority over gravity. Another notable death was that of Vincenzo Viviani, Galileo's last direct disciple and collaborator.

1653 (3.5 ¢)

Around the middle of the seventeenth century, anatomists inspired by William Harvey's teaching (1628, 3.75 ¢) about the circulation of the blood were investigating systems of vessels in various animals. Olof Rudbeck, son of the most influential Swedish cleric of the time, scooped the others. While not yet 20, he detected the lymphatic vessels proceeding from the liver in calves, sheep, dogs

and cats. By 1651 he had worked out the structure of the lymphatic system and its connection with the lymph glands. He took time out to write on the circulation of the blood, so that only this year can anatomists celebrate the 350th anniversary of his pamphlet *Nova exercitatio anatomica, exhibens ductus hepaticos aquosos*.

1603 (4.0 ¢)

Staying with the circulation of the blood, we recommend for notice Girolamo Fabricio's (b. 1533, 4.7 ¢) *De Venarum Ostiis*, the description of the valves of the veins on which Harvey drew when beginning his studies of the circulation. In almost the same line, Santorio Santorio, like his friends Fabricio and Galileo a professor at the University of Padua, invented many instruments to measure bodily functions including, in or around 1603, a pendulum for determining pulse rates.

Neither Andrea Cesalpino, physician to Pope Clement VIII and author of an excellent botany text *De Plantis* (published in 1583, 4.2 ¢), nor William Gilbert, physician to Queen Elizabeth I and the most important early writer on magnetism, had a pulse to measure when Santorio's meters appeared. Nor did François Viète, a lawyer and author of the earliest work on symbolic algebra, and, in 1593 (4.1 ¢), of a volume of problems for algebraists, *Zeteticorum Libri Quinque*.

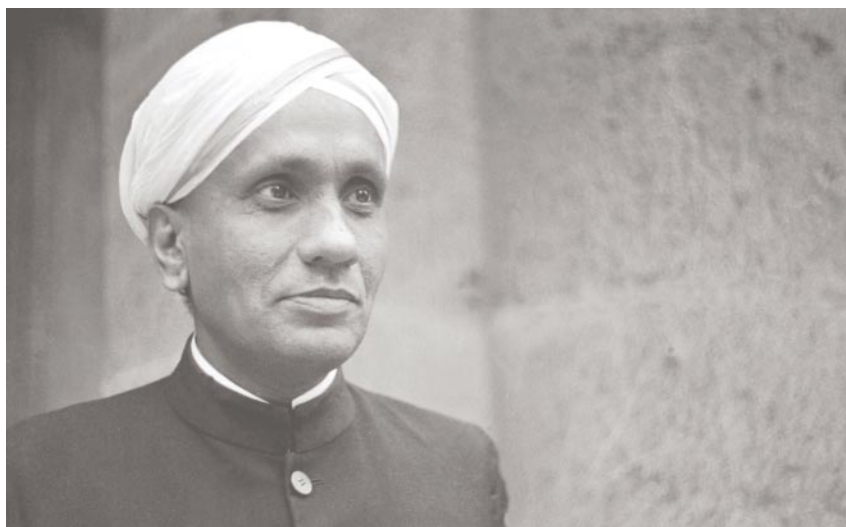
On the plus side, the Accademia dei Lincei, perhaps the earliest significant group of natural philosophers (Galileo belonged to it), was founded; Johann Bayer published his beautiful star atlas, *Uranometria*, which introduced the convention of designating the stars by constellation and by Greek letters according to their apparent brightness; and Pietro Cataldi found the sixth and seventh perfect numbers, 8,859,869,056 and 137,438,691,328.

1553 (4.5 ¢)

Michael Servetus, one-time printer, self-trained physician and unfortunate amateur theologian, published a theological treatise, *Christianismi Restituto*, in 1553. This



Shining star: the constellation of Cygnus from Johann Bayer's atlas *Uranometria*.



Visible progress: in 1928, Chandrasekhara Venkata Raman showed the particle-like nature of light.

contained the first printed account of the minor circulation, but also much that was anathema, notably denial of the doctrine of the trinity. For this Servetus went to the stake in 1553, in Calvin's Geneva.

Girolamo Fracastoro, a notable physician and author of the famous poem *Syphilis, sive morbus gallicus*, and Erasmus Reinhold, author of the astronomical Prutenic tables that reduced Copernicus's theories to art, also died 450 years ago, but not as violently as Servetus.

1403 (6.0 ¢)

Our sources say that in this year the Chinese published, in three copies, an encyclopaedia in 22,937 volumes. We feel more confident in reporting that in 1403 Venice established a quarantine against the Black Death, and that Cardinal Johannes Bessarion was born in Trebizond. The cardinal, who left Constantinople for Italy before it was too late, brought with him his library, which contained a Greek manuscript of Ptolemy's *Almagest* that he put at the disposal of European astronomers.

1253 (7.5 ¢)

Three-quarters of a millennium ago, give or take a year, Robert de Sorbon received permission from the King of France to found a college in the Latin Quarter of Paris; Robert Grosseteste, bishop, mathematician and natural philosopher, died; and the decimal system may have been introduced into England by John of Holywood (Johannes de Sacrobosco), if he did not die, as some say he did, in 1244.

1953 (0.5 ¢)

Had it concerned anything less than the bricks of life, the discovery of the structure of DNA might not have won our coveted award of anniversary of the year. Even within its own year it faced strong competition.

We are not thinking of that now inescapable necessity, the microwave oven, inspired by the melting of a chocolate bar in the pocket of Percy L. Spencer as he passed a source of rays at Raytheon Corporation. Nor have we in mind the fruition of another invention inspired by the contemplation of doubtful foodstuffs, in this case American beer; that is, Donald Glaser's (\$1960) bubble chamber. Or the first broadcast of commercial colour television in the United States. Nor even Alfred C. Kinsey's revelations of the sexual behaviour of American women.

Even more momentous, 1953 was the year the Earth began to slip. Maurice Ewing and Bruce Heezen, following up an observation by their cartographer colleague Marie Tharp, recognized the existence of deep valleys running along the axes of mid-ocean ridges. The finding, which they announced in 1956, became a major ingredient in the theory of plate tectonics.

Another strange and important fact that came to light in 1953 was strangeness itself. In order to classify certain 'elementary'

particles, Murray Gell-Mann (\$1969) introduced a new quantum number, "strangeness", to delay the decays of kaons and hyperons that lived longer than theoreticians had previously allowed.

Also in 1953 it was discovered that mice that smoked cigarettes, or, at least, that dined on cigarette tars, developed malignant tumours. Evarts A. Graham and Ernest L. Wynder supplied this fact, which now seems strange only in the stubbornness of those who refused to acknowledge its bearings on humans.

Charles Hard Townes' (\$1964) invention of the maser in 1953 prompted the invention of the laser. The myriad applications of the latter, from operating theatres to supermarket check-outs, make the laser another worthy runner-up to the deciphering of the double helix.

1953 ± 25

Those who believe in progress might be shocked that the harvest of deep discoveries in 1928 exceeded that in 1978. In the earlier year, Alexander Fleming (\$1945) discovered the antibiotic properties of penicillin mould; Paul Adrien Maurice Dirac (\$1933) fielded the equation that unified quantum mechanics with spin and relativity, and that contained an apparent blemish later interpreted as a description of antiparticles; Chandrasekhara Venkata Raman (\$1930) showed that visible light has particle-like properties; and George Gamow (no prize, why not?) calculated that subatomic particles could tunnel through barriers they could not climb.

And what can be said of 1978? There was achievement in the very large and the very small. The space probe Pioneer Venus showed that the goddess of love has an atmosphere consisting almost entirely of carbon dioxide, and James W. Christy and Robert S. Harrington found that Pluto has a moon that weighs a tenth as much as it does. On the other end of the scale, Frederick Sanger (\$1958 and \$1980) and colleagues at Cambridge completed the first DNA sequence of an entire genome — that of bacteriophage ϕ X174 — and Robert A. Weinberg demonstrated the existence of oncogenes.

On the consumer front, 1978 offered the first disk drive for personal computers, by Apple; the first test-tube baby, by Robert G. Edwards and Patrick C. Steptoe; and the banning of chlorofluorocarbons as propellants in hair spray, by the US government. In 1928 we were given bubble gum, by Walter Diemer, and the answer to King Gillette (1903) — Jacob Schlick's electric razor. ■

J. L. Heilbron is at Worcester College, Oxford OX1 2HB, UK. W. F. Bynum is at the Wellcome Trust Centre for the History of Medicine, University College London, 24 Eversholt Street, London NW1 1AD, UK.



Louise Brown, the first test-tube baby, born 1978.

Two minds

The Vogts sought to tie 'psychic' functions to specific regions of the brain.

Cécile and Oskar Vogt: The Visionaries of Modern Neuroscience

by Igor Klatzo (with Gabriele Zu Rhein)
Springer Verlag; 2002. 130 pp. £76

Edward G. Jones

They met in 1897. Cécile Mugnier was a tall, broad savoyarde, slow-speaking but witty nonetheless. Against the wishes of her family she had come to Paris to study medicine. At the age of 23, while assistant to the neurologist Pierre Marie, she met Oskar Vogt, a short, dapper north German with a strong streak of contentiousness. He had learned neuroanatomy under Paul Flechsig at Leipzig, psychiatry with Otto Binswanger at Jena and hypnotism with August Forel at Zürich. He was already an accomplished exponent of the use of hypnosis in the treatment of neurosis and anxiety, and had come to Paris to study human brain anatomy under Jules Dejerine. Cécile and Oskar married in Berlin in 1899, and there commenced a 60-year scientific partnership known to posterity simply as 'the Vogts'.

On moving to Berlin, accompanied by 30 human brains donated by Marie, they established a private neurobiological centre, perhaps the first of its kind, and probably the first use of the term 'neurobiology'. Against institutional resistance, and only after the intervention of the armaments manufacturer F. A. Krupp, whom Oskar Vogt had befriended some years earlier, the centre was incorporated into Berlin University. In 1914 it became the Kaiser Wilhelm Institut für Hirnforschung, which in 1931 moved to extensive premises in the Berlin suburb of Buch.

The philosophy of the institute was directed towards the localization of 'psychic' functions in the brain through the use of refined anatomical methods. But there was also an abiding interest in extending these studies to reveal the basis for the superior capacities of artists, scientists, statesmen and others with 'élite' brains. Cécile seems to have had the better eye for neuroanatomy, whereas Oskar possessed the ability to transform observations drawn from anatomical studies, from the results of electrical stimulation in animals and from human neuropathology, into principles of brain organization and function.

Through the doors of the institute passed many who were to become leaders in their fields throughout the German-speaking lands, Poland and Russia. From its beginning, the institute was divided into sections devoted to anatomy, histology, brain stimu-



Inseparable: Cécile (right) and Oskar Vogt were married in 1899 and worked together for 60 years.

lation, physiology, genetics, clinical research, psychology and chemistry, each headed by an individual scientist. The most distinguished of these were to be the chief assistant Korbinian Brodmann, the histologist Max Bielschowsky, and the Vogts' elder daughter and neurochemist, Marthé. The work of visiting scientists, including M. Friedemann, Max Lewandowsky, Theodor Mauss, Rolf Hassler, the two Roses, Maximilian and Jerzy, from Poland, and J. L. Pines, and collaborations with other German scientists such as Alois Alzheimer, conferred even greater stature on the institute. There, the science of cyto- and myeloarchitectonics was born at the hands of Brodmann and the Vogts. Cécile Vogt defined the subcortical connections of the thalamus and basal ganglia and the pathology of the striatum, Bielschowsky discovered his silver stain, and Hassler described the loss of neurons from the substantia nigra that is the hallmark of Parkinson's disease.

As an old-style social democrat who had enjoyed strong associations with Soviet scientists and physicians, Oskar Vogt was not popular with the Nazi régime. For a time, his friendship with the Krupp family gave him protection, and his pugnacity in the face of ever-mounting pressure from the local authorities protected the institute and his place in it. Eventually, however, in 1936 he was eased out of the directorship, as much by age as by political considerations.

With the help of the Krupps, a new institute housing some of the invaluable collection of human brains from Berlin was established at Neustadt in the Black Forest, and there the Vogts saw out the Second World War in relative peace.

The house journal, *Journal für Psychologie und Neurologie*, which had carried most of the major neuroanatomical papers of a whole era, saw its final issue come out there, containing the last significant paper by the Vogts, on the connections of the human thalamus and basal ganglia. Oskar devoted himself largely to speculations about how genetic and evolutionary alterations in cytoarchitectonic units in cortical regions might confer vulnerability to stroke and other pathological insults. The institute stumbled on for a decade or so after Oskar's death in 1959 and Cécile's departure to join Marthé in Cambridge, but it was eventually wound down. The Berlin and Neustadt brain archives had by now been reunited and were transferred to Düsseldorf, where they have proved a useful resource for modern investigators.

Since the war, the Vogts have suffered from a perennially bad press. Their commitment to breaking up the 50 or so areas defined by Brodmann in the human cerebral cortex into finer and finer subdivisions (eventually reaching more than 200) did not sit well with the prevailing mood of modern neuroscience. Oskar's propensity for seeing

cellular evidence of 'élitense' in the brains of accomplished individuals did not help.

After Lenin's death in 1924, Oskar was called to Moscow to help set up the Moscow Brain Research Institute, which was to be devoted to the study of Lenin's brain. Lenin's brain does not seem to have been preserved very well and this, along with recognition that it was remarkably atrophic and possibly even syphilitic, precluded significant investigation. A few blocks from the parietal cortex, however, found their way into Vogt's hands and to Berlin. On examining sections from them, and apparently noticing an unusual number of large pyramidal cells in the superficial layers of this cortex, Vogt declared Lenin an *Associationsathlete*, with a pronounced ability to form connections in the brain. This may have reflected considerable percipience in recognizing, long before the conclusive demonstration in the late 1970s, that the large layer III pyramids are the origins of long-range corticocortical connections, or it may have been little more than a tongue-in-cheek concession to the Soviet authorities who had paid Oskar handsomely. But it did nothing to help the Vogts' reputation.

In the early 1990s, the Vogts suffered further ignominy in Tilman Spengler's *Lenin's Brain* (translated by Shaun Whiteside; Hamish Hamilton, 1993), an irreverent, and at times viciously funny, fictional send-up of the German neurological establishment during the Wilhelmine, Weimar and early Nazi periods. The revelation recently that some of the brains of schizophrenics in the Vogts' collection may have come from inmates of the concentration camps has cast a further pall.

Igor Klatzo, who spent three years working with the ageing Vogts at Neustadt-Schwartzwald in 1946–48, has written a personal tribute to his mentors in which he highlights those personal qualities that are unrevealed by documentation of their considerable scientific achievements. There is, unfortunately, not much discussion of the scientific contributions of the Berlin institute, which still await definitive analysis. However, all of the data that Spengler had previously used so incisively in his novel are presented here. Oskar's often fraught relations with Brodmann, and especially with Bielschowsky, are brought out as never before, and Klatzo's work will undoubtedly serve as a rich source of other anecdotes for the neuroscience community. Its most interesting feature, however, is its account of rural Germany getting back on its feet after the war, and of the camaraderie that the author and other young scientists experienced in the company of the Vogts and their remaining staff during this period. ■

Edward G. Jones is at the Center for Neuroscience, University of California at Davis, Davis, California 95616, USA.

Going back to the root

Foundations of Tropical Forest Biology: Classic Papers with Commentaries

edited by Robin L. Chazdon &

T. C. Whitmore

University of Chicago Press: 2002. 862 pp.

\$95, £66.50 (hbk) \$35/£24.50 (pbk)

Roderick J. Zagat

This book is intended as a reminder to those tropical biologists who, always looking up in their efforts to advance knowledge, forget to look down at the giants on whose shoulders they stand. This anthology of classic papers on tropical-forest biology assembles a comprehensive group of giants, including some that one never realized one was standing on.

Robin Chazdon and the late Tim Whitmore have put together facsimile excerpts of 56 seminal papers that mark milestones in our understanding of tropical biology. It is too easy to fall into "sarcastic if not acrimonious criticism both of omission and of commission", as the natural historian William Beebe put it, quoted here in the introduction. But it does seem that to be considered a foundation of tropical biology it helps to have written in English, rather than French, Spanish or Japanese.

The papers are categorized in 12 sections covering general areas such as evolution, diversity, species composition and ecosystem ecology. This raises the question of what constitutes 'tropical biology'. Are biological processes in the tropics fundamentally different from those elsewhere, or is it simply that the particular environment and history of the tropics allows these processes to produce the broadest diversity, the strangest adaptations and the most coevolved inter-specific interactions? The great majority of authors represented in this volume seem to be inspired by the wide diversity of tropical forests and contribute, from one angle or another, insights into the evolution and maintenance of high species richness in the tropics. This focus on the outcome — amazing diversity — rather than on the underlying biological processes *per se*, probably defines tropical biology. And it helps to explain why this book gives comparatively little attention to subjects that biologists deal with, such as ecophysiology or genetics, even in the tropics.

Each section starts with an introductory review by a recognized expert in the field, which puts the excerpted papers into the context of the past development of the concept. The papers are then discussed for their role in that development, and how the issues they raised have served to inspire and guide subsequent

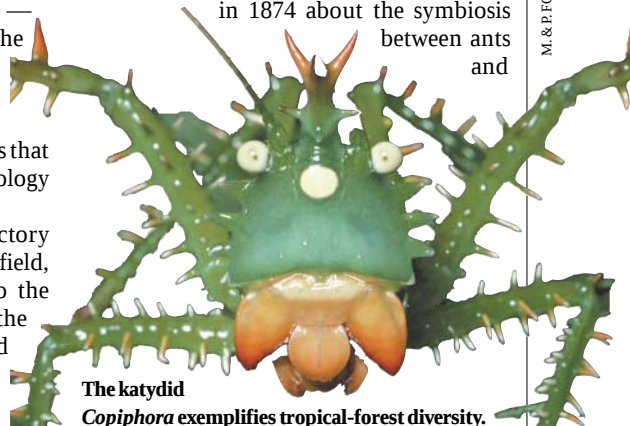
research. Several of these introductions provide excellent mini reviews of their topic.

The papers reprinted date from 1814 to 1987, with a strong emphasis on the 1960s and 1970s, but the section introductions cover the most recent advances in knowledge. This mix makes the book something of a hybrid between anthology and textbook. The anthology aims to highlight well-written historical papers that have made significant contributions to the development of knowledge. This is partly at odds with the duty of a textbook to be complete and describe the state of the art, when modern findings may have added to or even overtaken older research. An anthology can limit itself to the aims of entertainment and illustration. The editors could have been bolder in dismissing influential but bone-dry technical papers, such as the study by Raven and Axelroth on biogeography and continental movements, in favour of papers that add flavour to the scientific facts.

With the advantage of hindsight, the editors have selected historical papers that proved to be milestones along a wide and straight road towards increased understanding, rather than just some of the confusing signposts that mark the narrow and winding road to our current knowledge. Not every influential paper has survived the test of time. Similarly, the artificial cut-off date inevitably dismisses more recent crucial advances, even if they are more significant in the progress of knowledge than the ones reproduced. For example, it would be hard to find a researcher studying tree species diversity in tropical forests who is not influenced by the work of Stephen Hubbell, but none of his papers are included.

Most of these papers were written at a time when research in the tropics was hampered even more than it is today by poor accessibility and difficult working conditions. One is awestruck by the keen observation, clear thinking and artful pen displayed by many of the writers. Today, when many biologists seem to be more preoccupied with statistical than observational power, it is refreshing and stimulating to read the observations by Thomas Belt

in 1874 about the symbiosis between ants and



The katydid *Copiphora* exemplifies tropical-forest diversity.

M. & P. FOGDEN/CORBIS

plants, to mention just one example. If this book stimulates present-day researchers to pursue the same combination of art and science, it will have more than accomplished its objective. ■

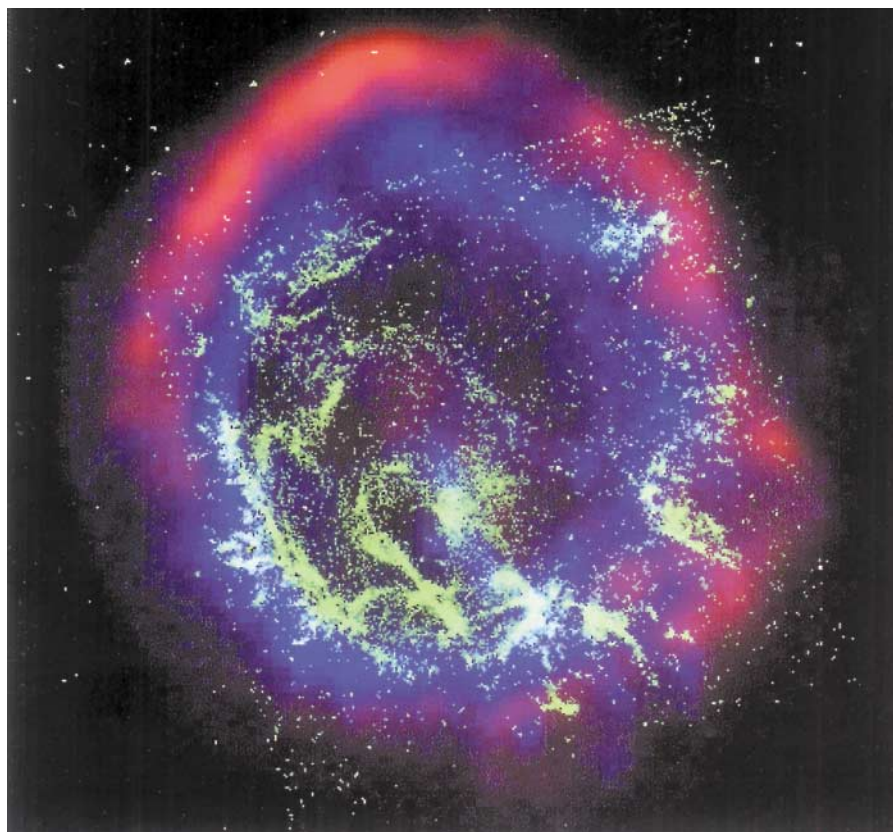
Roderick J. Zagt is at the Nationaal Herbarium Nederland, Utrecht University Branch, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands.

The turbulent cosmos

**The Restless Universe:
Understanding X-ray Astronomy in
the Age of *Chandra* and *Newton***
by Eric M. Schlegel
Oxford University Press: 2002. 228 pp. \$30
Luigi Piro

The award of part of this year's Nobel Prize in physics to Riccardo Giacconi for his role in the founding of X-ray astronomy has raised interest in a field that, in the past 40 years, has uncovered new and extreme phenomena in the cosmos. In the narrow bandwidth to which our eyes are sensitive, the sky seems fixed, dominated by constant and static sources, mostly stars and galaxies. But, equipped with their X-ray gear, astronomers have uncovered a very different and violent Universe, filled with pulsating, bursting, exploding sources, where the laws of physics are pushed to the extreme. The exotic population of the X-ray sky includes cosmic accelerators of all dimensions propelling particles close to the speed of light; neutron stars so dense that a teaspoonful of material would weigh as much as the Roman Colosseum on Earth; black holes orbiting around other stars and erupting blobs at relativistic speeds; and monster black holes lying in the centre of galaxies (including our own Milky Way).

X-rays also give us a direct view of the most abundant form of luminous matter in the Universe: the hot gas that sits in the gravity well of clusters of galaxies, the largest gravitationally bound structures in the Universe. Its temperature is so high (10^6 – 10^7 K) that this gas is visible only at X-ray wavelengths. At optical wavelengths, the cluster appears as a concentration of galaxies in the sky, but these galaxies are just little specks compared with the mass of hot glowing gas seen in X-ray images. Yet even this does not account for the total mass of the cluster, which is dominated by dark matter, the primary constituent of the mass of the Universe. The study of the hot X-ray gas should give researchers clues to the origin of this elusive dark matter. The properties of the hot gas are also affected by the matter ejected by the galaxies of the cluster, in particular by the heavy elements (metals) that are synthesized



X-ray data (blue) complement optical (green) and radio data (red) in this image of a supernova remnant.

in the final stage of a star's life through supernova explosions. The formation and evolution of stars, galaxies, clusters of galaxies and the whole Universe are therefore deeply connected, and X-ray observations provide a powerful probe into this cosmic web.

Eric Schlegel's *The Restless Universe* is a popular-science book that tells how X-ray astronomers are investigating the cosmos. Because our atmosphere blocks X-rays completely, instruments have to be located on satellites. Two important X-ray satellites are now orbiting Earth — NASA's *Chandra* and the European Space Agency's *Newton*. Schlegel, an astrophysicist at the Harvard-Smithsonian Center for Astrophysics, who has worked on the *Chandra* programme for several years, describes the challenges that X-ray astronomers faced in building this observatory and launching it into space. But the book goes far beyond this story. It gives the layman the background to comprehend both the present and future of X-ray astronomy. After a brief history of the field, the central part of the book covers the methods by which astrophysicists obtain X-ray images of sources in the sky, study their variations in time, measure their spectra, and interpret this information, deriving insight into the nature of the sources.

The working principles of the increasingly sophisticated instruments developed by researchers since the early days of X-ray astronomy are clearly recounted, using analogies with everyday life. And, for some

topics, the reader can easily track the advances of knowledge derived by the sharpening of techniques in imaging, timing and spectroscopy. Schlegel often widens this discussion to encompass observations at other wavelengths, and lets the reader appreciate the deeper insight that modern research in astrophysics gains by bringing together data from wavelengths from radio to gamma-rays. He then shows what astrophysics is all about: a leading edge of physics where the Universe is the laboratory in which physical laws under extreme conditions can be probed. The author concludes by describing some of the most challenging future experiments.

This guiding of the reader from techniques and experiments to science is the book's main strength but, almost inevitably, it results in a rather uneven and fragmentary presentation of the scientific panorama. Also, missing from the book are several of the most recent *Chandra* and *Newton* results (obtained after the book was written) and some of the major achievements of previous X-ray satellites. But these are minor quibbles. The book provides a highly readable account of the exploration of the Universe through X-rays, providing a solid foundation for those who wish to move on to read more about this exciting field of human knowledge. ■

Luigi Piro is in the Istituto Astrofisica Spaziale e Fisica Cosmica, CNR, Via Fosso Cavaliere, 00133 Rome, Italy.

Nature's insight into what makes scientists tick.

Summarize yourself in the form of a title of a paper in Nature.

Longitudinal single case study documents instance of a scientist 'making it' notwithstanding apparent lack of key skills.

What was your first experiment as a child?

I built many contraptions, including model airplanes that actually flew; in fact, I was very good with my hands. But it was books that peopled my world, even when I was small.

Who has been the most important mentor in your career?

My German adviser Gotthilf Hempel. He was politically conservative and me insufferably leftist (to the point of never calling him 'Professor'), but he supported me through thick and thin. He was a true *Doktorvater*.

Whose graduate student would you most like to have been (historical impossibility notwithstanding)?

Charles Darwin. In fact, we've just finished a book together — *Darwin's Fishes* — to which he contributed about a third of the text.

What single event changed your career path?

The sudden realization, while on a research trawler in the Java Sea in 1975, that we would never be able to make sense of the hundreds of species of multicoloured fish wiggling on our sorting deck using the approaches applied in Europe for studying cod or plaice, and which had formed the core of my studies as a fisheries scientist.

What book has been most influential in your scientific career?

Leftist political tracts swirling about in the 1960s, which convinced me that I should study something that would 'serve the people', rather than indulge in my literary, historical and philosophical interests. As it turned out, I was able to do it all.

What gives you the most job satisfaction now?

Interacting with the creative, hardworking colleagues I have coaxed into joining the Sea Around Us project (see www.saup.fisheries.ubc.ca).

What are your major frustrations?

I failed to convince anyone of the usefulness of the neat theory I developed in my dissertation of 1979. I showed that oxygen availability to tissues is usually the limiting factor in the growth of fish and water-breathing invertebrates, which explains the key features of their growth, reproduction and distribution much more elegantly than the hypotheses that still prevail today. Twenty years of preaching and writing on this have gotten me nowhere, although I think this is the best work I ever did.



Daniel Pauly

Daniel Pauly is professor of fisheries at the Fisheries Centre, Faculty of Graduate Studies, University of British Columbia, Vancouver, Canada. He has no hobbies to speak of, and he reads when he does not write.

What literary character would you employ as a postdoc?

Sherlock Holmes. I would put him in charge of that part of our project devoted to tracking, quantifying and mapping illegal fisheries catches.

What's your favourite conference destination, and why?

Downtown Vancouver, because I can go there without flying and without being randomly searched at every airport.

What was the worst/most memorable comment you ever received from a referee?

"It may be true in the Tropics, but not here!" I can't remember where this came from, but Alan Longhurst and I followed up on some of its implications in *Ecology of Tropical Oceans*, a book we published in 1987.

What book is currently on your bedside table?

Steven Pinker's *The Blank Slate*. I had serious doubts before I started this book that the human brain starts as a *tabula rasa*, which is what we all believed in the 1960s. Pinker showed that I can abandon this erroneous belief and still be a good person.

What music heads the playlist in your car or lab?

Aretha Franklin's.

What's the best piece of advice you've ever received?

"Any scientist who wants to make important

discoveries must study an important problem" — Peter Medawar in *Advice to a Young Scientist*.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Circe.

Where and when would you most liked to have lived or worked?

Subtract a few decades and I, as a black person, could not have been an academic anywhere. So I much prefer the present.

What one thing would you rescue from your burning laboratory?

Nothing. Let it all burn.

What do you most dislike about having research published?

Finding that I could now do the same work much better, but the publication holds it in the past, and that is where people find it.

The Internet is the bane of scientists' lives because...

People expect prompt answers to their e-mails, and you spend the best part of your day working for them. If you don't, they think you have become too full of yourself.

What would you have become, if not a scientist?

Another scientist. Seriously. This is the best.

What single discovery, invention or innovation would most improve your life?

An intelligent, voice-activated spreadsheet for the office, and an exercise machine that does for me the exercise I should be doing.

Name one extravagance you can now get away with because of your eminence.

Not knowing or not being able to do something, saying so loud and clear, and having people still think I do.

What music would you have played at your funeral?

Aretha Franklin's *R.E.S.P.E.C.T.*

How would you like to be remembered?

As the one who showed that the effect of fisheries on marine life is equivalent to that of a large meteor strike on terrestrial life.

What's just around the corner?

A representation of global ocean ecosystems that we immodestly call 'World Model', contrasting the structure of present marine ecosystems, and their fish biomass, with those of the 1950s. This should help to settle the issue of fisheries impacts, as well as providing a solid framework for discussions about future global fish supply as required, for example, by the United Nations-sponsored Millennium Ecosystem Assessment. ■

A tangled problem

William R. Taylor and Kuang Lin

From our experiences with shoelaces, we tend to think that tying a knot requires both dexterity and intent. As proteins lack both of these qualities, it was assumed for many years that there would be few, if any, knots found in protein chains. But perhaps shoelaces are not a good model for proteins. Before the Christmas decorations are packed away, it is possible to find a slightly better model in the strings of beads that are sometimes draped over the festive branches.

To conduct an unbiased experiment, take a metre length of these (around 150 beads, which represents a small protein), cover the end beads in Blu-Tack (adhesive putty) and 'pour' the beads from hand to hand. When the two end beads eventually stick together, check for knots. In an unscientifically small sample of 20 tries, one-quarter of our 'proteins' were knotted. From this result, one might now wonder why more proteins do not contain knots; but, until relatively recently, there were almost no protein chains that contained a knot that would not be laughed at by any qualified boy scout. Most of these had one end poking through a loop by only a few 'beads' (amino-acid residues).

Given that we now know the structures of around 2,000 different proteins, why are there so few knots? A likely explanation is that proteins are not free-flowing strings of beads, but rather are 'sticky'. (Dedicated experimentalists should now repeat the knot experiment, using beads coated in honey.) Interactions within the chain are therefore predominantly local, and few open loops will be formed for the protein's termini to pass through. However, a few years ago a knot in quite a different league was noticed in a protein; this knot had over 200 amino-acid residues on one side and 70 on the other. More recently, another respectable knot has appeared with 30 residues on its shorter side.

Unlike beads, real proteins do not (normally) have their termini joined. This presents a technical problem, as knots are only properly (mathematically) defined in circular strings. However, one common definition of a knot is "a loop in a string that tightens when pulled". This can be applied to a protein by repeatedly smoothing (averaging consecutive triples of points along the chain) while keeping the two termini fixed in place and seeing if a straight line is obtained. Using this method, the ends can also be progressively trimmed to find the exact location of the knotted core. Protein knots can then be quantified by the number of residues on either side of them, and a useful distinction can be made between 'deep' knots and 'shallow' ones, with the latter having less than 20 residues on their shorter side.

The 'true' or 'topological' knots considered above are defined by the path of the backbone chain alone. But there is another, more common, knot found in proteins that is created by crosslinks between parts of the chain. As the bonds involved are covalent (usually disulphide links), these knots are best referred to as 'covalent knots'. Unlike topological knots, there is no mystery about how covalent ones might form: they simply require amino acids with sulphide groups (cystine residues) to lie close enough to become crosslinked either during or after folding. Like disulphide bonds in general, their function may be to give extra stability to the fold. Might topological knots have a similar function, or are they just 'harmless' tangles that have arisen accidentally? Intriguingly, both of the known examples of deep knots occur in the catalytic domains of their proteins, with one even running through the active site. It is difficult to imagine anything in the structure of a knot that could not just as easily be constructed by an unknotted piece of protein chain. Any advantage from their presence must therefore derive from an

Protein knots

Do knots give extra stability to protein folds or are they just 'harmless' tangles that have arisen accidentally?

indirect or entropic (ordering) effect, such as reducing thermal motion in a knotted active site or allowing large motions only in a restricted segment of chain.

Another twist to the protein-knot story is provided by the structure of a SET-domain protein. The 'knotted' region in this protein fold is neither a topological nor a covalent knot. Instead, it consists of a loop held by hydrogen bonds that 'traps' the carboxy-terminal part of the chain. Perhaps our concept of a protein knot should be broadened to include these 'pseudo-knots' that are formed by hydrogen bonding as well as covalent links. The definition of a protein knot then becomes a matter of energetics: how many hydrogen bonds are needed? Or perhaps we need a continuum of crosslinks from covalent bonds, through hydrogen bonds, to van der Waals forces?

With the deluge of data anticipated from the various structural-genomics programmes currently being undertaken, we may soon have a large enough collection of folds to get a better idea of how frequent and important all forms of knots are to protein structure and function. If some of the ideas discussed above are correct, most will be of the shallow kind, and the few deep knots may require unusual folding mechanisms to account for them. The analysis of these should provide useful insight into how proteins fold. If knots are selected for some advantageous reason, then it might be expected that they will provide greater advantages in thermophilic bacteria. As the structures of proteins from both mesophilic and thermophilic organisms are being determined (the latest knot is from a thermophile), their comparison should provide a natural test-bed for this idea. ■

William R. Taylor and Kuang Lin are in the Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

FURTHER READING

Protein Data Bank <http://www.rcsb.org./pdb/> (PDB accession numbers for knotted proteins: 1QMG, 1YVE, 1IPA, 1GOZ, 1K3R, 1MT6, 1MVH, 1H3I, 1ML9, 1MLY).
Taylor, W. R. *Nature* **406**, 916–919 (2000).
Taylor, W. R., May, A. C. W., Brown, N. P. & Aszodi, A. *Rep. Prog. Phys.* **64**, 517–590 (2001).
Taylor, W. R., Xiao, B., Gamblin, S. J. & Lin, K. *Comp. Biol. Chem.* (in the press).
Mansfield, M. L. *Nature Struct. Biol.* **1**, 213–214 (1994).



M. GEORGE

The lobster navigators

Thomas Alerstam

When experimentally displaced in geomagnetic space, spiny lobsters act as if to make their way home. This is a fascinating case of navigation by an invertebrate using a magnetic map sense.

Sometimes an idea is so irresistible that it is revisited over and over again. One such is the notion that Earth's magnetic field provides guidance for animal orientation and even true navigation. This topic has a long history, and on page 60 of this issue Boles and Lohmann¹ present the latest turn of events. From the results of elegant displacement experiments in both geographical and geomagnetic space, they conclude that the impressive homing capacity of spiny lobsters is probably based on a map sense involving geomagnetic cues. This is the first time that such true navigational capability has been plausibly ascribed to an invertebrate.

During the first half of the nineteenth century, great advances in the understanding of magnetism were made in basic physics and through expeditions to the north and south magnetic poles. Duly inspired, in 1855 Middendorff² suggested that migrating birds behave like compass needles, unfailingly heading towards magnetic north in spring, guided by an inner magnetic sense caused (he conjectured) by induced electric currents in their bodies. Debate about the homing instinct in this journal in 1873 and 1874 included Darwin's crucial point³ that

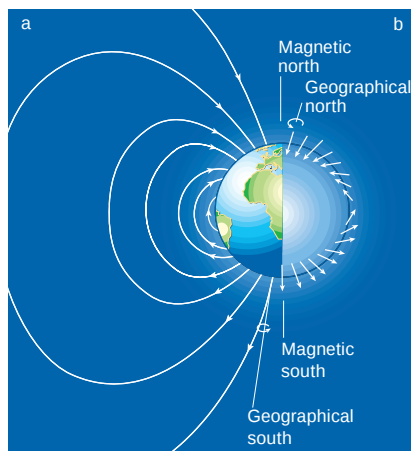


Figure 1 Position determination and Earth's magnetic field. **a**, The overall field lines of the magnetic field. **b**, The resulting magnetic inclination, or dip angle, and field strength at Earth's surface (indicated by arrow angle and length) vary in a consistent way. In theory these parameters make it possible for an animal to determine its position from magnetic coordinates. Magnetic declination (or variation) is the angular difference between magnetic north and geographical north. So it cannot be estimated from magnetic information alone, which excludes it as a possible map coordinate in experiments where animals have magnetic information only.

a compass sense alone is not enough to explain the spectacular navigation feats among animals: "Even if we grant to animals a sense of the points of the compass, of which there is no evidence, how can we account, for instance, for the turtles ... on the shores of the Isle of Ascension, finding their way to that speck of land in the midst of the Atlantic Ocean?"

Influenced by this debate, Viguié⁴, a

French zoologist working in Algeria, proposed in 1882 that the geomagnetic field is not only the basis of a compass sense but also furnishes animals with information about their position. He believed that the coordinates provided by magnetic inclination (dip angle), field intensity and/or declination would be sufficient for possession of a complete map sense (Fig. 1). This hypothesis was revived several times in the twentieth century^{5–7}, one version⁷ invoking coordinate information input from the Coriolis force as well. But each time the idea met with ephemeral support.

Only much later, by about 1970, did behavioural experiments with migratory birds and homing pigeons provide compelling evidence of the existence of a compass sense based on magnetoreception⁸. Since then, such a sense has been demonstrated in a wide range of animals — molluscs, crustaceans, insects, fishes, amphibians, reptiles and mammals⁸ — and there has been the discovery of biogenic magnetite as a possible basis for magnetoreception⁹. Hence the recent revival of interest in the topic.

All of which brings us to the spiny lobster, *Panulirus argus*, which lives in coastal waters in the Gulf of Mexico and the Caribbean. It undertakes seasonal migrations of up to 200 km, and has a remarkable capacity for homing even if displaced to deep waters. Some years ago, experiments carried out by Lohmann and colleagues¹⁰, employing underwater magnetic coils, showed that the lobsters have a magnetic compass sense that allows them to walk on the sea floor, in darkness and without hydrodynamic

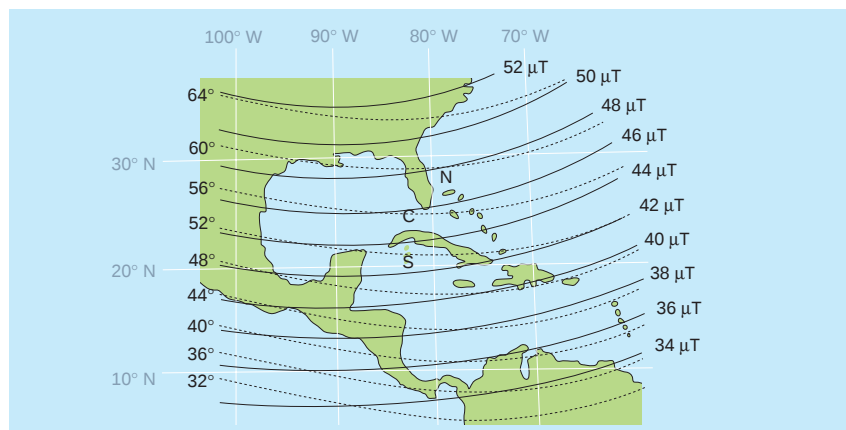


Figure 2 Magnetic inclination and total field intensity in the distribution range of the spiny lobster, *Panulirus argus*. Boles and Lohmann¹ found that lobsters at capture site C exposed to magnetic conditions typical of site N (+3.5° in inclination and +2.6 μT (microtesla) in total intensity relative to C) and site S (−4.4° in inclination and −2.5 μT in intensity), oriented themselves towards magnetic south and north, respectively. That is, they acted as if homing following real geographical displacement. Further experiments¹ indicated possible magnetic navigation over much shorter displacement distances (12–18 km), east–west as well as north–south. Magnetic navigation from east to west may be more difficult than from north to south, because of the small angle between different magnetic coordinates. And the magnetic parameters in this region are changing at about $-0.05^\circ \text{ yr}^{-1}$ and $-0.1 \mu\text{T yr}^{-1}$, corresponding to an annual displacement of gridlines of about 5 km and 20 km, respectively. How the lobsters cope with both sets of circumstances merits further investigation.

cues, along a constant course determined by the horizontal component of the geomagnetic field.

Boles and Lohmann¹ have now turned to analysis of the lobsters' homing capacity, carrying out a series of experiments in which lobsters were displaced 12–37 km from their capture site. The lobsters were deprived of visual or chemical cues (and probably also lacked inertial cues) during transport from the capture site to the test site. They were likewise denied such cues during the tests themselves. But they nonetheless showed an impressive ability to start walking in the direction of the capture site — that is, back home. This held true even when magnets were used during transport to the test site to disrupt magnetic compass information. Apparently the lobsters have a sense of position relative to home that is based on information picked up at the test site.

The final step in Boles and Lohmann's study was to keep the lobsters at the same geographical position, but to displace them in geomagnetic space by exposing them to magnetic conditions of inclination and field intensity that simulated positions to the north or the south of the capture site (Fig. 2). The lobsters responded as one would expect if they were determining their position from a geomagnetic map.

There is increasing experimental evidence of magnetic positioning and navigation in birds^{8,11}, sea turtles¹² and newts¹³. This study with the spiny lobster adds the first invertebrate to that list. But plenty of questions remain. Here are four.

First, does the geomagnetic field provide a basis for uni- or bicoordinate navigation — that is, does it allow positioning in only one dimension or (if two parameters form a useful grid) in two dimensions? Second, the lines of equal dip angle and field intensity run almost (but not exactly) parallel east–west in the spiny lobster's range (Fig. 2), and one would think that this circumstance would make magnetic navigation more difficult along this axis than latitudinally. But the lobsters showed no difficulty in detecting east–west natural displacements. Do they use geomagnetic cues in this situation and, if so, what geomagnetic elements might be involved?

Third, does the magnetic landscape in the study area (the local pattern of different magnetic gradients) satisfy the requirements for unambiguous determination of position? Fine-scale charting of the magnetic field will be needed to infer how magnetic navigation might work over such short distances. And fourth, the magnetic field is slowly changing in a way that corresponds to several kilometres of annual displacement of magnetic parameters in the study region: how do the lobsters keep their magnetic map information up to date?

Finally there is the more general question

of the sensory mechanisms that underlie the magnetic compass and navigational abilities. Progress has been made in studies of magnetoreceptor systems based on both light-dependent chemical (radical pair) processes and magnetite¹⁴. But all of the mechanisms that have been proposed remain hypothetical — this, above all, is the matter to which we would like to see some answers. ■

Thomas Alerstam is in the Department of Animal Ecology, Ecology Building, Lund University, SE-22362 Lund, Sweden.

e-mail: Thomas.Alerstam@ekol.lu.se

1. Boles, L. C. & Lohmann, K. J. *Nature* **421**, 60–63 (2003).

2. Middendorff, A. v. *Mém. Acad. Sci. Saint-Petersbourg Ser. 6* **8**, 1–143 (1855).
3. Darwin, C. *Nature* **7**, 360 (1873).
4. Viguer, C. *Rev. Phil. France Etr.* **14**, 1–36 (1882).
5. Maurain, C. *La Nature (Paris)* **51**, 232–238 (1923).
6. Stresemann, E. *Ardea* **24**, 213–226 (1935).
7. Yeagley, H. L. J. *Appl. Phys.* **18**, 1035–1063 (1947).
8. Wiltschko, R. & Wiltschko, W. *Magnetic Orientation in Animals* (Springer, Berlin, 1995).
9. Kirschvink, J. L., Jones, D. S. & MacFadden, B. J. (eds) *Magnetite Biomineralization and Magnetoreception in Animals* (Plenum, New York, 1985).
10. Lohmann, K. J. *et al. J. Exp. Biol.* **198**, 2041–2048 (1995).
11. Fransson, T. *et al. Nature* **414**, 35–36 (2001).
12. Lohmann, K. J., Cain, S. D., Dodge, S. A. & Lohmann, C. M. F. *Science* **294**, 364–366 (2001).
13. Fischer, J. H., Freaake, M. J., Borland, S. C. & Phillips, J. B. *Anim. Behav.* **62**, 1–10 (2001).
14. Lohmann, K. J. & Johnsen, S. *Trends Neurosci.* **23**, 153–159 (2000).

Quantum computing

Putting it into practice

Jonathan Jones

Will quantum information theory ever lead to practical quantum information technologies? At a conference reviewing the advances of the past two years, delegates looked to the future with cautious optimism.

Quantum information processing¹ (QIP) combines the counterintuitive theory of quantum mechanics with the practical field of information technology. Instead of worrying about the strange behaviour of quantum systems, researchers have harnessed this 'quantum weirdness' to perform apparently impossible types of information processing. Although the first ideas were formulated more than 30 years ago², it is only in the past decade that serious proposals for quantum technologies have been made. These involve individual quantum systems, such as atoms, ions, nuclei or photons, which can be manipulated to store and process information — but which must be carefully isolated from their surroundings, as any uncontrolled interactions, known as decoherence³, will destroy the fragile quantum effects.

One of the first methods used to investigate QIP involves trapping small numbers of ions with electromagnetic fields and manipulating them with laser pulses. As reported at a recent conference*, and on page 48 of this issue⁴, ion-trap methods have now been used to implement the simplest possible quantum computation, the Deutsch–Jozsa algorithm (R. Blatt, Univ. Innsbruck). Although the basic elements involved have been demonstrated before^{5,6}, this is the first unambiguous demonstration of quantum computation. The Deutsch–Jozsa algorithm is most accurately described in terms of function evaluation, where it provides a method for determining the parity of a binary function using only one function evaluation. An easier analogy is to imagine

tossing a coin. A normal coin has two sides, heads and tails, but a trick coin may show the same pattern on both sides. To recognize a trick coin, it would be necessary to look at first one, then the other side of the coin. But, by the Deutsch–Jozsa algorithm, in a quantum device this check could be made in effectively a single glance.

The original ion-trap approach to quantum computing, in which all the ions are held in a single trap, unfortunately cannot be extended to build large-scale devices. An alternative is to use a network of memory traps, in which ions are stored when not in use, and interaction traps, where computations occur. It has now been shown that ions can be moved between two traps while maintaining their quantum states (D. Wineland, NIST Boulder).

Ions interact strongly with electromagnetic fields and with one another, making them easy to trap and manipulate, but also rendering them vulnerable to decoherence. Atoms interact less strongly, so decoherence should be less of a problem. A lattice of traps can be built using intersecting laser beams to produce a standing light wave. Inside the trap, atoms seek the regions of lowest light intensity, which occur at the nodes of the standing wave. These traps, however, are very shallow, making it easy for atoms to escape.

The solution is to start with ultracold atoms, which have so little thermal energy that they remain trapped in this optical lattice (W. Phillips, NIST Gaithersburg). Ideally there should be only one atom at each lattice site⁷. The atoms can be made to interact by altering the properties of the laser beams, causing them to move. Although it is not yet possible to control the atoms

*Practical Realizations of Quantum Information Processing, The Royal Society, London, UK, 13–14 November 2002.

individually, detailed control of the whole lattice is possible (I. Bloch, MPI für Quantenoptik, Garching), allowing complex entangled states of many atoms to be created.

Entanglement, in which the properties of different atoms are inextricably intertwined, is a key quantum phenomenon that underlies quantum information processing. Entangled atoms could be used to implement certain quantum algorithms (I. Cirac, MPI für Quantenoptik, Garching). In particular, it should be possible to use this system to simulate the properties of other, less well-controlled, quantum systems.

Trapped atoms and ions can be studied today, but many researchers believe that the future of QIP lies with solid-state devices (A. Briggs, Univ. Oxford). One early proposal⁸ that generated a great deal of excitement uses the spin of phosphorus nuclei implanted in a silicon matrix to store the quantum information. The nuclei are controlled and observed through the behaviour of their surrounding electrons.

Building such a device would be extremely difficult, involving placing individual atoms at precisely known locations inside a silicon chip and then building tiny electrodes and transistors around them. These ideas have been demonstrated (R. Clark, Univ. New South Wales) and a device with two phosphorus atoms has been constructed. Early results suggest that these two atoms can be controlled and observed. Another proposal based on superconducting quantum interference devices (SQUIDS) has also made substantial progress (G. Wendin, Chalmers Univ., Göteborg). Several working single-SQUID devices have been constructed, and attempts to couple two SQUIDS are under way.

In addition to the new results announced at the meeting, several themes for future work emerged. Although many simple demonstrations of QIP have been achieved, only one of these, quantum cryptography (J. Rarity, QinetiQ), could really be described as practical⁹. But before we can make existing toy systems do something useful, we will have to greatly improve the precision of quantum logic gates (J. Jones, Univ. Oxford) and of sources and detectors (I. Walmsley, Univ. Oxford).

It has long been noted that there seems to be a close connection between the controllable interactions needed to implement quantum logic, and the uncontrollable interactions that cause decoherence. Is this simply an artefact of the systems studied so far, or does it reflect some underlying principle? In fact, whenever two quantum objects interact with one another they must also interact with their environment (A. Fisher, Univ. College London). The strengths of these desirable and undesirable interactions are inevitably related, and real quantum

computers will need this ratio to be as large as possible. To achieve this will require an extremely careful choice of environment.

Will these ideas lead to a practical quantum computer? The progress so far is remarkable, but the difficulties can hardly be overstated. Whether or not we eventually get there, we will certainly see some interesting scenery on the way.

Jonathan Jones is at the Centre for Quantum

Computation, Clarendon Laboratory, University of

Oxford, Parks Road, Oxford OX1 3PU, UK.

e-mail: jonathan.jones@qubit.org

1. Bennett, C. H. & DiVincenzo, D. P. *Nature* **404**, 247–255 (2000).
2. Wiesner, S. *SIGACT News* **15**, 78–88 (1983).
3. DiVincenzo, D. & Terhal, B. *Phys. World* **11** (3), 53–57 (1998).
4. Gulde, S. *et al.* *Nature* **421**, 48–50 (2003).
5. Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. *Phys. Rev. Lett.* **72**, 4714–4717 (1995).
6. Jones, J. A. & Mosca, M. J. *Chem. Phys.* **109**, 1648–1653 (1998).
7. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. *Nature* **415**, 39–44 (2002).
8. Kane, B. E. *Nature* **393**, 133–137 (1998).
9. Kurtsiefer, C. *et al.* *Nature* **419**, 450 (2002).

Neurobiology

The importance of depression

Charles F. Stevens

We, and other animals, can generally pinpoint the source of a sound in space regardless of how loud it is. A study involving experimentation and computer modelling reveals how our brains perform this clever task.

Information flows from one nerve cell to the next at highly specialized points of contact called synapses. All synapses exhibit depression¹ — a decrease in the strength of the connection that occurs after rapid and repeated use — and this might seem like a bad idea, a design fault that prevents synapses from keeping up with the demands placed on them. But on page 66 of this issue, Cook and colleagues² show that depression is a feature, not a bug, in auditory synaptic transmission. In fact, synaptic depression is used to compute sound levels and to correct neuronal signals, so that the brain's representation of the source of a sound in space is not confounded by the intensity of the sound.

Comprehending the contribution made by Cook *et al.*² depends on understanding one of the main ways in which we determine the spatial source of sounds³. You may have noticed that when you hear a series of noises whose source you cannot locate — a bird chirping, for example — you automatically turn your head as you listen. This is because you determine the location of the sound by using the tiny differences in the time it takes for the sounds to reach your two ears. And you turn your head to maximize this difference. If you are facing the source of a sound, the signal arrives simultaneously at both ears, but if the sound originates from, say, your left side, the sound pressure wave arrives a little earlier at your left ear than at your right.

The maximum difference in arrival times depends on head size, but is less than a couple of milliseconds even for elephants, and less than half a millisecond for us. However, we have clever neural machinery to compute the sound's spatial source from this slight difference. This computation takes place in the medial olivary nucleus in the brains of mammals (in birds, the nucleus laminaris). Neurons stimulated directly by each ear feed

into this brain region. There, the tiny time differences in the firing of the ear-stimulated neurons are turned into a 'place code', whereby interaural time differences are represented by which olivary neurons are responding and by how much³.

The trick the brain uses to change time differences to a place code involves delaying the neural signals from one ear relative to the other by different amounts for different olivary neurons, and then having these neurons fire an impulse only if the signals arrive simultaneously from both ears. That is, the medial olivary nucleus uses delay lines (conduction delays over nerve 'wires', or axons, of different lengths) and coincidence detection (firing of olivary neurons only when impulses arrive from both ears simultaneously) to construct a place code for submillisecond interaural time differences. This coincidence detection involves some clever circuitry⁴.

In addition, impulses produced by the ear-stimulated neurons in response to a high-frequency tone are 'time-locked' to the tone, in the sense that each impulse is produced at a particular phase of the sound pressure wave. If this time-locking did not occur, the coincidence-detection scheme described above would not work, because the brain needs precise information about the relative arrival times of sounds at the two ears. But a nerve impulse is not generated on each cycle of the sound pressure wave; rather, many cycles are skipped in any particular nerve fibre, so that only by looking across a population of inputs can the brain know when each wave arrived. As sound intensity increases, however, nerves fire on more and more of the cycles. So, the average number of nerve impulses per second depends on sound intensity, whereas the timing of each impulse, relative to the phase of the sound wave, is independent of intensity.

The fact that higher-intensity sounds

produce more nerve impulses causes a potentially serious problem for the neuronal mechanisms used to locate the source of the sound. Each olivary neuron doing the coincidence detection receives inputs from many neurons stimulated by each ear, so that the system still works even when any particular axon fails to produce a nerve impulse on a specific cycle of the sound wave. But if too many axons from one ear produce impulses on a particular cycle — something that could happen with loud sounds — then the coincidence of the many inputs from just that ear would be capable of producing an output in the medial olivary nucleus (or the nucleus laminaris), and the localization mechanism would break down. Behaviourally, though, we and other animals can pinpoint a sound's spatial source pretty accurately irrespective of its intensity. How do the coincidence detectors keep from being overwhelmed by loud sounds? This is the question that Cook *et al.*² asked.

Because the problem with loud sounds is that they cause nerve impulses to be produced on too many cycles of the sound pressure wave, Cook *et al.* reasoned that an easy way for the brain to compensate for this would be to decrease the amount of neurotransmitter released in response to each impulse. Neurotransmitters are the chemicals that pass signals at synapses from one neuron (here, an ear-stimulated neuron) to the next (an olivary neuron). In other words, if the synapses were correctly depressed with use, this could compensate for the greater number of impulses.

To test this idea, Cook *et al.* characterized the properties of depression at the input synapses of the nucleus laminaris in chicks, and found that these synapses do indeed decrease in strength with use. But can this depression actually compensate for the too-large inputs that result from loud sounds? To find out, the authors constructed a biophysically realistic computer model of neurons from the nucleus laminaris, and incorporated the synaptic depression as found in their experiments. They could then examine the effect of synaptic depression on coincidence detection by simply turning depression on or off in the model. Indeed, Cook *et al.* found that coincidence detection in their model works well, irrespective of sound loudness, when synaptic depression is present, but not when it is absent. Synaptic depression is a feature of neural computation, not a bug.

Short-term synaptic plasticity — the increase and decrease of synaptic strength that depends on the history of synaptic use — is a prominent feature of all synapses¹, and can easily cause several-fold changes in synaptic strength. Although various authors have speculated that short-term plasticity provides a functionally important dynamic filter for arriving nerve impulses (see the first four citations in ref. 2), testing this notion

is difficult and the earlier work was mainly speculative. Cook *et al.*² have provided the first documented use for such plasticity. Now we need to understand the mechanisms through which the correct properties of synaptic plasticity are set up in development and then adjusted throughout life. ■

Charles F. Stevens is at the Molecular Neurobiology Laboratory at the Salk Institute, and the Howard

Hughes Medical Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.

e-mail: stevens@salk.edu

1. Zucker, R. S. *Annu. Rev. Neurosci.* **12**, 13–31 (1989).
2. Cook, D. L., Schwandt, P. C., Grande, L. A. & Spain, W. J. *Nature* **421**, 66–70 (2003).
3. Goldberg, J. M. & Brown, P. B. *J. Neurophysiol.* **32**, 613–636 (1969).
4. Brand, A., Behrend, O., Marquardt, T., McAlpine, D. & Groth, B. *Nature* **417**, 543–547 (2002).

Planetary science

Sodium at Io

Donald M. Hunten

It has been known for some years that Jupiter's satellite Io has sodium as a component of its atmosphere. The source, it now seems, is sodium chloride emitted by volcanoes on Io's surface.

Io, the innermost of Jupiter's four Galilean satellites, is anomalous in various respects. One curiosity is its atmosphere, and the 'plasma torus' of ionized matter through which the planet orbits. On page 45 of this issue, Lellouch *et al.*¹ describe observations that will help settle thinking about a puzzling constituent of the atmosphere and torus, and its origins.

The true dimension of Io's strangeness first became apparent only in 1974, when R. A. Brown identified the intense emission

of yellow 'sodium D lines' in spectroscopic studies of Io. It was soon established that these emissions come from an extended region — the 'neutral torus' — that is much larger than the visual disk of Io and which shares its orbit around Jupiter^{2,3}. Two years later came the discovery that this region is also occupied by Io's plasma torus, which consists of ionized material that has been created by the impact of electrons on neutral molecules and atoms. The satellite's orbit is embedded in the torus, which has turned



Figure 1 Images of Io. The main picture, in true colour, was taken by the Galileo Orbiter in 1999. The inset, showing Jupiter with Io at the lower left, is an image from the Voyager 1 mission (1979). Io is about 3,600 km in diameter.

NASA

out to be the source of most of the ions in Jupiter's magnetosphere.

Ground-based measurements provided further understanding of the torus. But it was spacecraft encounters with Io (Fig. 1) that provided the next surprises. Data from Voyager 1's 1979 encounter not only revealed intense emissions in the far-ultraviolet part of the spectrum, primarily due to sulphur and oxygen ions in several states of ionization, but also identified widespread volcanism and geyser activity on Io's surface. The source of the sulphur and oxygen ions was quickly shown to be sulphur dioxide; and, simultaneously, it was realized that the volcanic and geyser activity was due to intense internal heating of Io, a result of the tidal flexing by Jupiter. Studies of the Galileo Orbiter data have established that some of the volcanism occurs at unusually high temperatures.

But what was the origin of the sulphur and oxygen in the neutral atmospheric clouds and plasma torus? The ions in the torus are forced by Jupiter's magnetic field to rotate at nearly the same rate as the planet, and therefore bombard Io's atmosphere and surface at energies of hundreds of electron volts. This was identified as the source of the sulphur and oxygen — through a process called sputtering, ion bombardment of SO₂ on Io's surface releases atoms and molecules or ions into the atmosphere and neutral torus; the atoms and molecules are then dissociated and ionized by energetic electrons in the plasma torus. But the nature of the compound constituting the sodium source remained a puzzle. Theoretical modelling had shown that it is probably sodium chloride, and this is now confirmed by Lellouch *et al.*¹, who report its detection in vapour form.

This part of the story begins two years ago with Küppers and Schneider's identification of chlorine ions in the plasma torus⁴. Their observation in the near infrared was quickly confirmed in the far ultraviolet⁵. On the theoretical side, Fegley and Zolotov⁶ extended their earlier predictions of the gases and vapours that would be expected to be emitted by volcanic magmas on Io. To their assumed magma composition, taken to occur at high temperatures (1,000–2,000 K), they added sodium, potassium and chlorine in cosmic abundances. This chemical model produced copious NaCl vapour (along with sodium atoms, potassium chloride, SO₂ and other sulphur compounds).

Using the 30-m IRAM (Institut de Radio-Astronomie Millimétrique) radio telescope, sited near Granada in Spain, Lellouch *et al.* have now identified two spectroscopic lines of NaCl vapour, at wavelengths of 1.3 mm and 2.1 mm, in Io's atmosphere. They also detect three lines of SO₂ and provide a tentative detection of KCl. The NaCl occurs only in patches in the atmosphere, consistent with there being just a few volcanic sources. An alternative, but much less likely, explanation

is that the NaCl stems from evaporation or sputtering from material condensed on SO₂ snowdrifts on Io's surface. Earlier observations of SO₂ with the same telescope show that SO₂ is also distributed patchily, and for this more volatile molecule both sources — volcanism and sputtering — are plausible.

The amount of Na and Cl in the torus, in various forms, has been calculated to be only some 2% of the total composition. The NaCl supply rate found by Lellouch *et al.* is too large to be consistent with this figure, but their estimate includes some uncertain parameters and the agreement appears to be satisfactory. Thus, 28 years after Brown's discovery of sodium at Io, we may be

close to understanding how it reaches the atmosphere and torus. ■

Donald M. Hunten is in the Lunar and Planetary Laboratory, and the Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721, USA.

e-mail: dhunten@pl.arizona.edu

1. Lellouch, E., Paubert, G., Moses, J. I., Schneider, N. M. & Strobel, D. F. *Nature* **421**, 45–47 (2003).
2. Brown, R. A., Pilcher, C. B. & Strobel, D. F. in *Physics of the Jovian Magnetosphere* (ed. Dessler, A. I.) 197–225 (Cambridge Univ. Press, 1983).
3. Spencer, J. R. & Schneider, N. M. *Annu. Rev. Earth Planet. Sci.* **24**, 125–190 (1966).
4. Küppers, M. & Schneider, N. M. *Geophys. Res. Lett.* **27**, 513–516 (2000).
5. Feldman, P. D. *et al.* *Astrophys. J.* **554**, L123–L126 (2001).
6. Fegley, B. Jr & Zolotov, M. Yu. *Icarus* **148**, 193–210 (2000).

Molecular evolution

Duplication, duplication

Axel Meyer

Duplicated genes are common in genomes, perhaps because they provide redundancy: if one copy is inactivated, the other can still work. A new study quantifies the effects of deleting 'singletons' and duplicated genes in yeast.

In fairy tales, things frequently come in twos: there are, for instance, two witches ruling over different parts of the land of Oz, two ugly sisters vying for the attention of Cinderella's prince, and so on and so on. And the phenomenon of duplication is not restricted to stories. In eukaryotes (loosely speaking, those organisms, such as humans, whose DNA is packaged into cell nuclei), genomes seem to be far from optimally designed, in that most stretches of DNA sequence do not code for proteins, and even those small portions that do are often duplicated. Why do organisms tolerate such apparent wastage? Gu and colleagues¹ tackle this question on page 63 of this issue, looking specifically at the effects of duplicated genes on the 'fitness' of individuals.

An important line of thinking about why duplicated genes might arise goes back 30 years to Susumo Ohno², who stated that "natural selection merely modified while redundancy created". Ohno reasoned that gene (and even genome) duplications are not a burden on the organism, but rather the raw material for evolutionary diversification — in other words, duplication allows new gene functions to evolve. One copy of a gene can carry out the original task while the duplicate becomes free to accumulate mutations, possibly developing new functions and allowing the big steps in evolution to occur. In today's era of wholesale genome sequencing, Ohno's hypothesis has gained many new adherents through the recognition that duplicate genes are abundant in most genomes and that significant portions of genomes are repeated. But, in general, the actual effects of 'singletons' and duplicated genes on evolutionary

fitness — that is, on roughly how well different individuals fare compared with others in terms of reproduction — have not been well studied at the whole-genome level.

On the other side of the coin, gene duplicates appear to have another important function: they can buffer the genome against environmental perturbations and mutations, because if one copy of the gene is somehow inactivated, another with the same or a similar function can be used instead. Such genetic redundancy is a headache for researchers trying to determine the role of a particular gene, because the standard technique of knocking out that gene in an organism might not have a noticeable effect, thanks to functional substitution by the duplicate. Gu *et al.*¹ shed new light on this issue.

The authors studied the budding yeast *Saccharomyces cerevisiae*, which has become the workhorse of the functional-genomics community³ since the publication of its genome sequence six years ago. In a previous high-throughput study⁴, nearly every one of the organism's 6,357 genes was knocked out, one at a time. The effects were then investigated under five different culture conditions, in which components of the culture medium that affect fermentation and respiration were varied. The fitness of each mutant yeast strain was measured by its ability to proliferate, relative to non-mutated strains.

Gu *et al.* have taken advantage of these data to determine the effects on fitness of duplicated and single-copy genes. Using a computational method, the authors started by searching for yeast genes that have a high degree of sequence similarity — that is, those that are alignable over at least 50% of their length. The

reasonable assumption behind this approach is that sequence similarity between two or more genes suggests that they are copies of an initially single gene. In this way, Gu *et al.* determined that about a quarter of all *S. cerevisiae* genes exist as duplicates. The degree of sequence similarity is presumed to provide some insight into when the genes were duplicated and how similar their functions are. Immediately after duplication, two genes will have identical sequences and functions. With time, their sequences and therefore, presumably, their functions will diverge more and more, or one copy will stop working completely (becoming a pseudogene).

Next, Gu *et al.* reanalysed the previously published data on *S. cerevisiae* fitness⁴ in light of their knowledge about the number of gene duplicates. They first show that knocking out singletons generally reduces fitness more severely than deleting one gene of a pair of duplicates. For instance, knocking out only 12.4% of genes that are part of a pair, compared to 29% of singletons, has a lethal effect — a significant difference. Moreover, knocking out 64.3% of genes that have a duplicate, but only 39.5% of singletons, has no or only a weak effect on fitness. Can these results be explained by functional compensation between duplicated genes (at least under these conditions)? If so, one might predict that deleting one duplicated gene of a pair would have a similar effect on fitness to deleting the other gene, but that deleting two randomly selected genes would have more pronounced differences in terms of fitness effects. This is exactly what Gu *et al.* find.

One would also predict that duplicated genes that are highly similar in sequence should be better at compensating for each other than duplicates whose sequences (and, presumably, functions) have diverged further. Gu *et al.* confirm this prediction, too. On average, however, even highly diverged duplicates still have weaker fitness effects when deleted than singletons, suggesting that some degree of functional compensation persists even here. The authors also looked at the effects of the level of expression of gene duplicates, using data from microarray experiments⁵. They found that negative fitness effects were more pronounced when a duplicated gene with a higher level of expression than its partner was deleted.

All of this suggests that gene duplication provides a means of preserving function; even when two copies of a gene have diverged widely, they can still substitute for each other functionally to some degree. This, together with the fact that many genes and gene networks are similar in evolutionarily diverse species, hints that maybe Ohno was wrong after all. Are duplicated genes the stuff of developmental stability and of conservation of function rather than evolutionary innovation? If so, how did the diversity of life around us appear?

The answer may lie in part with such factors as the evolution of novel gene-regulatory sequences, the generation of alternative protein products from a single gene, and the recruitment of duplicated genes into diverse functions and into many different networks of interactions. The importance of these mechanisms in relation to gene and genome duplication remains unclear. But gene duplication is an important phenomenon to understand. The discovery of many duplicated genes and parts of genomes has been an unexpected but interesting by-product of genome-sequencing projects. And recent comparative genomic analyses found a duplication rate of about 1% per gene per million years^{6,7}, which is quite high on evolutionary timescales.

We are only now beginning to comprehend just how malleable genomes are, and also how resilient they are in the face of so much genetic perturbation; for instance, rearrangements and duplications of chromosomal segments are also commonplace^{8,9}. Gu *et al.*¹ have provided the first estimate (23–59%) of the contribution of duplicated genes to genetic robustness. This may be one

reason why duplicated genes do not diverge to produce pseudogenes, or 'die', as quickly or as often as had been predicted on the basis of population-genetics theory¹⁰. I would guess that the existence of multiple gene functions and their recruitment into novel gene networks provide another explanation. But more needs to be learned about the evolution of gene networks, through comparisons of complete genome sequences and through further functional-genomic analyses, before this question can be answered. ■

Axel Meyer is in the Department of Biology, University of Konstanz, 78457 Konstanz, Germany.
e-mail: axel.meyer@uni-konstanz.de

1. Gu, Z. *et al.* *Nature* **421**, 63–66 (2003).
2. Ohno, S. *Evolution by Gene and Genome Duplication* (Springer, Berlin, 1970).
3. Grünfelder, B. & Winzler, E. A. *Nature Rev. Genet.* **3**, 653–661 (2002).
4. Steinmetz, L. M. *Nature Genet.* **31**, 400–404 (2002).
5. Causton, H. C. *et al.* *Mol. Biol. Cell* **12**, 323–337 (2001).
6. Lynch, M. & Conery, L. C. *Science* **290**, 1151–1155 (2000).
7. Lynch, M. & Conery, L. C. in *Genome Evolution: Gene and Genome Duplications and the Origin of Novel Gene Functions* (eds Meyer, A. & Van de Peer, Y.) (Kluwer, Dordrecht, in the press).
8. Prince, V. E. & Pickett, F. B. *Nature Rev. Genet.* **3**, 827–837 (2002).
9. Betrán, E. & Long, M. *Genetica* **115**, 65–80 (2002).
10. Cooke, J. *et al.* *Trends Genet.* **13**, 360–364 (1997).

Condensed-matter physics

Two bodies are better than one

Bernard Barbara

Single-molecule magnets can change their spin states through quantum tunnelling. A more complete picture of the interactions occurring in a system of such magnets must include two-body transitions.

To avoid the complexity of the macroscopic world and get at the fundamental mechanisms of nature, physicists prefer to study ensembles of simple, identical objects, such as quantum dots or single molecules. In time, these objects may become the elementary parts that are assembled into functionalized devices: as Richard Feynman said, physical laws do not imply any limitation on our ability to construct and assemble objects at the atomic scale.

Although single atoms can be handled relatively easily, manipulation of the quantum properties of spin systems must be achieved if the potential of quantum-spin devices is to be realized. As well as having applications in 'spintronics', spin manipulation is a bridge to understanding a basic problem in quantum mechanics — how the quantum properties of spins are affected by their environment¹.

Another step towards understanding a quantum spin in its environment has been taken by Wernsdorfer *et al.*² with their study of single-molecule magnets (SMMs), published in *Physical Review Letters*. SMMs carry a well-defined spin, *S*, and can be arranged in model systems so as to minimize the

magnetic interactions between them³. Spin reversal can then occur through the quantum process of tunnelling^{4,5}.

Typically, SMMs have a large uniaxial magnetic anisotropy — that is, the directions of their spins are oriented either one way or the other (usually represented as '+' and '−'). In their study, Wernsdorfer *et al.*² used an SMM with the chemical formula $\text{Mn}_4\text{O}_3(\text{OSiMe}_3)(\text{OAc})_3(\text{dbm})_3$ — known simply as Mn_4 — that, in its ground state, has spin $S = 9/2$. In a magnetic field, the energy levels associated with the different spin states of the SMM suffer 'Zeeman splitting' (Fig. 1a): the spin states split into positive and negative values, and the energy of the $-9/2$ state becomes higher than that of the $+9/2$ state, and similarly for the other spin states available to the SMM ($\pm 7/2$, $\pm 5/2$, $\pm 3/2$ and $\pm 1/2$, a total of $(2S + 1)$ energy states).

At low temperature, where thermal fluctuations are negligible, an SMM can reverse its spin (from a positive to a negative value, or vice versa) by tunnelling through the energy barrier between the potential wells of two spin states^{4,5}. For the conservation of energy to hold, the energy levels associated with each state must coincide. This tunnelling

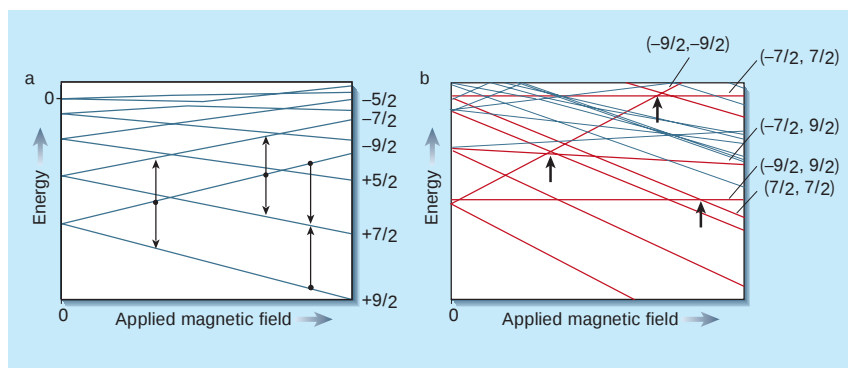


Figure 1 Spin states and energy levels. **a**, In a magnetic field, the energy levels in an Mn_4 single-molecule magnet (SMM) become split, and the splitting increases with the strength of the applied field. States with negative spin values have higher energy; states with positive spin values have lower energy. Two individual SMMs may make simultaneous transitions between states, as shown, that are equal and opposite in magnitude, and overall energy is conserved. **b**, The actual scheme of energy levels for a coupled two-molecule system is more complex. The transitions seen by Wernsdorfer *et al.*² are indicated. These 'spin-spin cross-relaxations' occur at the crossing of lines representing different spin states: because the energy levels of these states coincide, the SMMs can change state by quantum tunnelling. (Graphs derived from ref. 2.)

process is now familiar for single molecules, but Wernsdorfer *et al.*² saw evidence of more spin transitions that cannot be attributed to the usual single-molecule process. It seems that these extra transitions, which are much smaller than single-molecule transitions, result from simultaneous changes of state of interacting molecular spins.

In their crystalline samples of Mn_4 (between 10 and 500 μm in size), Wernsdorfer *et al.*² pinpointed three particular transitions corresponding to simultaneous changes in the spin of two SMMs: spins equal to $-9/2$ and $-9/2$ become $-7/2$ and $9/2$, or $-7/2$ and $7/2$; and spins $-9/2$ and $9/2$ become $7/2$ and $7/2$ (Fig. 1b). In each case, only one spin tunnels (changes its sign), while the other spin makes a transition within the same well (the value changes, but not the sign). This phenomenon is an example of spin-spin cross-relaxation (an effect originally described for a situation without an energy barrier⁶), and is called exchange-bias tunnelling, because the value of the magnetic field at which this tunnelling occurs is shifted to restore energy conservation. Simultaneous changes of state between interacting spins have already been seen^{7,8} in an ensemble of Ho^{3+} (holmium) ions diluted in a crystalline matrix of LiYF_4 . But in this case the effect was observed at relatively high temperatures, in the thermally activated tunnelling regime of Ho^{3+} ions. Wernsdorfer *et al.*² have now shown that exchange-bias tunnelling is also at work at low temperature in SMMs.

These additional transitions affect the broadening of spectral lines that is observed in SMMs. When an atom or molecule in a magnetic material changes energy state, a phonon is emitted whose energy corresponds to the size of the transition. But the

existence of extra exchange-bias tunnelling transitions (and related 'co-tunnelling' transitions, in which both spins change sign^{7,8}) means that the possible transitions for the system get closer and closer together, to the extent that the multi-spin transitions could eventually fill in the space between single-spin transitions. But, for co-tunnelling transitions, going from a single spin to two interacting spins means that there is a rapid decrease in the tunnelling probability: put simply, it's harder to change $2S$ to $-2S$ than to change S to $-S$. If the number of spins involved is increased further, the tunnelling probability decreases exponentially.

Wernsdorfer and colleagues' study of two-molecule interactions² has produced a more complete picture of a spin system. But further studies of multi-spin transitions in the tunnelling regime of both molecules and ions are needed if the consequences of spin manipulation are to be completely understood — and if networks of quantum spins are to be viable as the bits of a quantum computer. ■

Bernard Barbara is in the Laboratoire de Magnétisme Louis Néel, 25 Avenue des Martyrs, BP 166, 38042 Grenoble Cedex 09, France.
e-mail: barbara@grenoble.cnrs.fr

1. Prokof'ev, N. V. & Stamp, P. C. E. *J. Low Temp. Phys.* **104**, 143–210 (1996).
2. Wernsdorfer, W., Bhaduri, S., Tiron, R., Hendrickson, D. N. & Christou, G. *Phys. Rev. Lett.* **89**, 197201 (2002).
3. Sessoli, R. *et al.* *J. Am. Chem. Soc.* **115**, 1804–1816 (1993).
4. Thomas, L., Lionti, F., Ballou, R., Gatteschi, D., Sessoli, R. & Barbara, B. *Nature* **383**, 145–147 (1996).
5. Friedman, J. R., Sarachik, M. P., Tejada, J. & Ziolo, R. *Phys. Rev. Lett.* **76**, 3830–3833 (1996).
6. Bloembergen, N., Shapiro, S., Pershan, P. S. & Artman, J. O. *Phys. Rev.* **114**, 445–459 (1959).
7. Giraud, R., Wernsdorfer, W., Tkachuk, A. M., Mailly, D. & Barbara, B. *Phys. Rev. Lett.* **87**, 057203 (2001).
8. Giraud, R., Wernsdorfer, W., Tkachuk, A. M., Mailly, D. & Barbara, B. *J. Magn. Magn. Mater.* **242**, 1106–1108 (2002).



100 YEARS AGO

... the pamphlet entitled "Survey of the Sciences," which forms an appendix to a paper on University Development, is of especial importance at the present time, for we are glad to know that the belief that the weakness of our universities must lead to national weakness in several directions is growing with a rapidly accelerating pace. It may be long in this slow-moving country before the influence of Brain-power on history is recognised as fully as the influence of Sea-power has been, thanks to Captain Mahan, but undoubtedly it will be bad for our future if much more time is lost ... We are glad to see that the *Times*, in a sympathetic article, goes to the root of the matter in stating that "if the pocket of the millionaire is closed, the pocket of the nation must be opened." Our eleven universities are competing with 134 State and privately endowed in the United States and twenty-two State endowed in Germany. English private endowment is much less than 10 per cent. of the American endowment, and the German State gives to one university more than the British Government allows to all the universities and university colleges in England, Ireland, Scotland and Wales put together.
From *Nature* 1 January 1903.

50 YEARS AGO

On December 22, 1938, a fish of Crossopterygian type was taken by trawl-net, at a depth of about 40 fathoms some miles west of East London. The animal was unquestionably alive when caught. It was 1,500 mm in total length and weighed 127 lb. The colour was a bright metallic blue which faded to brown with preservation. Miss Courtenay-Latimer, curator of the East London Museum, took charge of the catch, and eventually Prof. J. L. B. Smith, now research professor of ichthyology in Rhodes University, Grahamstown, was able to examine it (*Nature*, 143, 455; 1939) ... Now comes a report of a further catch. This second fish was reported as having been caught on December 20 off the island of Anjouan in the Comoro group, two hundred miles west of Madagascar. It was reported to be about 5 ft. long and to weigh about 100 lb. On December 29, the fish was received by Prof. Smith, who had flown to the island to retrieve it and to take it to Rhodes University for further study.
From *Nature* 3 January 1953.

Obituary

Arthur T. Winfree (1942–2002)

High above the roiling waves on the NaPali Coast of Kauai Island, Hawaii, Art Winfree is kneeling down and photographing the lichens inscribed on the rocks along the Kalalau trail. He had passed the same way several years before and had documented the spiral patterns of lichen growth. These seemed similar to the spiral waves of activity that were now well known in certain types of chemical reactions and in the electrical rhythms of the heart. Could the lichens be growing with the same geometries, though on an infinitely slower timescale? Art Winfree's pioneering studies of the topological features of biological oscillations defined whole new areas of research. He died on 5 November 2002, struck down by an aggressive glioblastoma that had invaded his brain.

Winfree's scientific path was already set by the time he attended high school in Stanford, Connecticut. He figured out ways to blow huge soap bubbles, and for this feat he was one of the Westinghouse Science Talent Search finalists in 1961. He went on to study engineering physics at Cornell University, and then embarked on graduate studies of circadian rhythms — the roughly 24-hour oscillations of physiology and behaviour that most organisms display — at Princeton University.

Before completing his doctorate, Winfree was recruited to a faculty position in the Department of Theoretical Biology at the University of Chicago. There, he launched two different experimental efforts — to reset the circadian rhythms of fruitflies using short light flashes, and to describe the patterns of an oscillating chemical reaction, the Belousov–Zhabotinsky (BZ) reaction. Although these appear to be in completely different scientific domains, Winfree demonstrated theoretical links that led to testable predictions. To understand the connection, first think of a rotating disc, for example an old-fashioned long-playing record. At the very centre of the disc lies a fixed point that does not move at all. Winfree recognized that, in any mathematical equation describing a biological or chemical oscillator, there must also be a fixed point that does not lead to oscillation.

But how could this fixed point be observed experimentally? Well, there are lots of ways. For instance, if stimuli are delivered to the oscillator at different phases of the cycle, and the magnitude of resetting of the oscillator cycle is plotted



Scientist who shaped studies of biological and chemical oscillations

as a function of the timing of the stimuli, the graph will have different topological properties for small and large stimuli. Winfree demonstrated these features in studies of the circadian rhythms of fruitflies in the 1970s, using light pulses as the stimuli. Much later, Charles Czeisler and colleagues at Harvard Medical School showed that similar phenomena could be observed after perturbing the human sleep–wake cycle with light pulses.

For chemical reactions, if the fixed point could be trapped in a thin layer of a solution, spiral waves of chemical activity would be observed. In a landmark 1972 paper, Winfree rediscovered spiral waves in the BZ reaction (an earlier report of such waves by Anatol Zhabotinsky and Albert Zaikin was not well known outside the Soviet Union). Winfree's lectures on this work were enlivened by his distribution of the BZ reaction on filter papers to his audience.

At the same time that Winfree's studies were under way in Chicago, scientists in the former Soviet Union were expanding on a conjecture (by Norbert Wiener and Arturo Rosenblueth) that spiral waves might underlie fibrillation — a fatal heart rhythm. Recognizing the importance of this work, Winfree arranged for and edited a translation of a 1984 monograph by Vladimir Zykov that summarized the contributions of the Soviet school.

Winfree extended this work by showing how a stimulus delivered to the heart would lead to the initiation of fibrillation.

But to analyse the geometry of wave propagation in the pumping chambers of the heart, it would be necessary to consider waves in three dimensions.

Winfree introduced the concept of scroll waves to describe the three-dimensional rotating waves that, in cross-section, appear as a spiral. He now teamed up with Steven Strogatz and, in a beautiful series of papers in the 1980s, they classified the ways that scroll waves could be twisted, linked and knotted in three dimensions.

In contrast to the current scientific environment, which encourages large research grants and large teams, most of Winfree's work was done alone or in collaboration with one or two students or colleagues, and with comparatively small sums of money used in imaginative ways. This approach is evident in one of his last major studies, carried out at the University of Arizona, his academic home after 1989. The work involved the visualization of scroll rings in three-dimensional chemical reactions, and for this purpose Winfree built a three-dimensional optical tomograph that was cobbled together from miscellaneous bits of equipment.

Winfree was also generous with his ideas. His masterpiece *The Geometry of Biological Time* (first published in 1980), for instance, is a gold-mine of suggestions for new scientific directions, and a second edition (2001) identifies almost 200 research questions. Sending a manuscript to Winfree before publication would lead to a meticulous set of annotations, in green felt-tip pen, pointing out areas for further digging. He was scrupulous about crediting his colleagues, and his papers and books (which also include *The Timing of Biological Clocks*, 1986, and *When Time Breaks Down*, 1987) are notable for their comprehensive nature, containing hundreds of references.

From the growth of lichens on rocks, to the delicate patterns of chemical waves, to fatal cardiac arrhythmias, Art Winfree delighted in finding unexpected connections between geometry and dynamics. As researchers develop new ways to visualize the patterns of excitation in chemical systems and heart tissue, Winfree's predictions about the initiation, geometry and stability of spiral waves and scroll waves are being tested and are yielding new insights. His many friends, colleagues and members of his family rejoice in his contributions, and mourn the passing of this wise and gentle man.

Leon Glass

Leon Glass is in the Department of Physiology, 3655 Drummond Street, McGill University, Montreal, Quebec H3G 1Y6, Canada.
e-mail: glass@cnd.mcgill.ca

A catapult action for rapid limb protraction

Energy bursts from a horse's elastic biceps muscles provide power for a flat-out gallop.

Fast runners must be able to protract their limbs quickly in order to prepare for the next stance phase^{1–3}. This is particularly challenging for large animals as their limbs are long and their muscles contract slowly and have a low power output^{5,6}. Here we show that horses cannot achieve the high power output required for rapid limb protraction by simple muscle contraction and that they instead deploy an elastic biceps muscle to store and then release bursts of energy — this muscle's catapult action has an output that is comparable to over 100 times its mass of non-elastic muscle. Although grasshoppers and fleas are known to rely on a similar catapult mechanism for rapid acceleration^{7,8}, to our knowledge this has not been demonstrated before in larger animals.

The legs of large cursorial animals function like pogo sticks, storing and returning energy during running gaits. This is a spring–mass system, so energy storage and return occupy similar periods of time^{7,9}. In a catapult, energy is stored slowly with a large force but is released quickly to accelerate a small mass^{7,10}. This mechanism, however, requires a more sophisticated lever, or cam system, to exert sufficient force on the spring and then release it^{7,10}. Such a system exists in the horse (Fig. 1) — the forward movement of the trunk and the orientation of the ground–reaction force (GRF) during stance stretches the biceps muscle while the carpus is locked in extension; in late stance, the carpus buckles and releases the leg.

At 50% of stance (Fig. 2a), the GRF is flexing the shoulder joint (resisted by the biceps), flexing the elbow (resisted by the triceps and digital flexor muscles) and extending the metacarpophalangeal joint (resisted by the digital flexors and suspensory ligament). At 90% of stance (Fig. 2a), the GRF has moved further behind the shoulder and is now also behind the elbow joint. The GRF is therefore further flexing the shoulder and extending the elbow, both of which stretch the biceps.

Through stance, the carpus is extended by the GRF and the tendinous link of lacertus fibrosis/extensor carpi radialis, and is flexed by the digital flexors and palmar carpal ligament. When the GRF approaches the neutral axis of the carpal joint, it no longer balances the digital flexors' moment (this occurs at around 90% of stance). The carpus then buckles forwards, destabilizing the limb and releasing the catapult. The GRF and the digital flexors subsequently fold the collapsed joint and the limb is

accelerated forwards by recoil of the biceps. At 101% of stance (Fig. 2a), the biceps muscle is flexing the elbow and extending the shoulder.

We used force-plate and motion analysis to work out the limb–joint angles and GRF moment at the shoulder of thoroughbred horses during trotting at 3 m s^{−1} (n = 16). We determined biceps arm moment at the elbow and shoulder joints by radiography and dissection; tendon paths were roughly circular and were therefore assumed to be so. Eighty per cent of shoulder extensor moment was attributed to biceps and 20% to supraspinatus (apportioned from muscle architecture¹¹). Biceps length peaked at 91% of stance with a force of 11.4 kilonewtons (kN), and 74 joules (J) of elastic energy was stored. The biceps muscle has a substantial internal tendon (cross-sectional area, 200 mm²) which is parallel with two short-fibred (6 and 18 mm, respectively) pennate muscle heads¹².

Figure 2b shows the force–length relationship measured for biceps *in vitro* (n = 7) and *in vivo*. The similarity of the plots indicates that the muscle's properties *in vivo* are mainly passive. During galloping on a treadmill at 12 m s^{−1}, maximum biceps length was 12 mm greater than in trotting (at 100% of stance). Extrapolation of the curves shown in Fig. 2b by 12 mm gives a peak internal tendon strain of 7.5%, a force of 19.8 kN, and an elastic-energy storage of 261 J.



Figure 1 Good going: horses stretch and release their biceps in a gallop, creating 100 times the energy output of non-elastic muscle.

The energy released after the foot leaves the ground ('foot-off') was determined from biceps length and from the results shown in Fig. 2b, assuming that muscle fibres are isometric⁹ and that energy loss due to hysteresis is 7% (ref. 7). The sum of the kinetic and potential energy of the limb segments below the elbow (relative to the horse's centre of mass and segment height at foot-off, respectively) mirrored the biceps' elastic energy release, peaking at 33 J in a trot (Fig. 2c) and 120 J in a gallop. Some elastic energy was released before foot-off (Fig. 2b), and some energy accelerated the scapular and humeral segments and extrinsic muscle tissue.

The kinetic and potential energy of the limb continued to increase, but more slowly, after biceps energy/force had dropped to zero. This may indicate a shortening of the contractile component as it unloaded

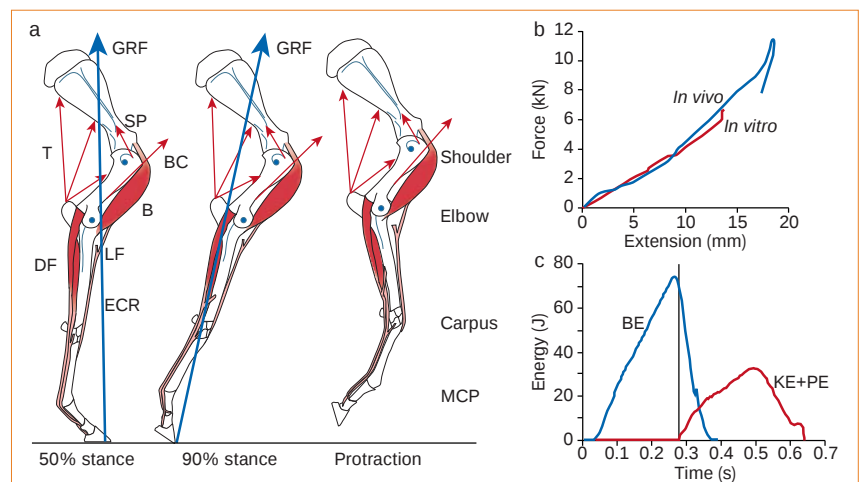


Figure 2 Mechanics of equine limb protraction. **a**, Equine forelimb at 50%, 90% and 101% (protraction) of stance, taken from our motion-analysis data. Biceps (B), brachiocephalicus (BC), supraspinatus (SP), triceps (T), lacertus fibrosis (LF), extensor carpi radialis (ECR) and digital flexor (DF) muscles are indicated. GRF, orientation of the ground–reaction force vector. MCP, metacarpophalangeal joint. **b**, Force–extension relationship for biceps *in vivo* in a trot (n = 16) and during tensile testing *in vitro* (n = 7). *In vivo* plot shows unloading as carpus and elbow flex before the foot is raised. **c**, Energetics of limb protraction. The foot makes contact at time zero and leaves the ground at the time indicated by the vertical line. Blue curve (BE) represents the elastic energy stored in the biceps through the gait cycle; red curve shows the sum of the kinetic and potential energy (KE + PE) of the radial, metacarpal and digital segments after foot-off.

(a common phenomenon) and/or work by the brachiocephalicus muscle. We devised a five-segment computer simulation of the equine limb that had biceps represented by a spring ($k = 700 \text{ kN m}^{-1}$) and recorded the joint angles at foot-off. The joint angular acceleration curves and segment trajectories generated by the simulation were very similar to those that we observed *in vivo* (see supplementary information).

The biceps released 243 J in 0.11 s in a gallop (2,200 W), which equates to a muscle with a peak power output of roughly 4,400 W (ref. 10). Equine muscle has a peak power output of about 90 W kg^{-1} ($V_{\text{max}} = 3.0$ lengths per second^{5,6}; $F_{\text{max}} = 0.3 \text{ MPa}$ (refs 5, 9, 10); $a/P_0 = 2.5$ (ref. 5)). A horse would therefore require 50 kg of non-elastic muscle to achieve the same power output as a 0.4-kg biceps muscle. The peak power output of brachiocephalicus (identified as the main forelimb protractor¹³) would be only 220 W. The catapult mechanism is therefore essential for rapid protraction of the equine limb, and it is likely that similar mechanisms exist in other animals.

Alan M. Wilson*†, Johanna C. Watson*,

Glen A. Lichtwark†

*Structure and Motion Laboratory, Department of Veterinary Basic Sciences, The Royal Veterinary College, Hatfield, Herts AL9 7TA, UK
e-mail: awilson@rvc.ac.uk

†Institute of Human Performance, University College London, Stanmore, Middlesex HA7 4LP, UK

1. Weyand, P. G., Sternlight, D. B., Bellizzi, M. J. & Wright, S. *J. Appl. Physiol.* **89**, 1991–1999 (2000).
2. Pratt, G. W. & O'Connor, J. T. *Am. J. Vet. Res.* **39**, 249–253 (1978).
3. Hutchinson, J. R. & Garcia, M. *Nature* **415**, 1018–1021 (2002).
4. Schmidt-Nielsen, K. *Scaling: Why is Animal Size so Important?* (Cambridge Univ. Press, Cambridge, 1984).
5. Woledge, R. C., Curtin, N. A. & Homsher, E. *Monogr. Physiol. Soc.* **41** (Academic, London, 1985).
6. Rome, L. C., Sosnicki, A. A. & Goble, D. O. *J. Physiol.* **431**, 173–185 (1990).
7. Alexander, R. McN. *Elastic Mechanisms in Animal Movement* (Cambridge Univ. Press, Cambridge, 1988).
8. Bennet Clark, H. C. *J. Exp. Biol.* **63**, 53–83 (1975).
9. Biewener, A. A. & Roberts, T. J. *Exp. Sport Sci. Rev.* **28**, 99–107 (2000).
10. Peplowski, M. M. & Marsh, R. L. *J. Exp. Biol.* **200**, 2861–2870 (1997).
11. Biewener, A. A. *Comp. Biochem. Physiol. B* **120**, 73–87 (1998).
12. Hermanson, J. W. & Hurley, K. J. *Acta Anat.* **137**, 146–156 (1990).
13. Dyce, K. M., Sack, W. O. & Wensing, C. J. G. (eds) *Textbook of Veterinary Anatomy* 2nd edn, 542–575 (Saunders, London, 1987).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Outer planets

Origins of atmospheric zonal winds

The origin of the strong zonal (east–west) jets at cloud-top level in the atmospheres of Jupiter and Saturn¹ is something of a mystery^{2,3}. Using an idealized two-dimensional, deep-turbulence model applied to a rapidly rotating planetary interior that is mixed by thermal convection, we are able to simulate the atmospheric multiple jets of these giant

planets, and show that deep origins may explain one of their distinguishing features — the prograde (westerly) equatorial jets.

This feature cannot be reproduced in conventional shallow-atmosphere models. A set of idealized simulations of single fluid-layer flows on a rotating sphere, starting from a random initial condition⁴, explains both the magnitude and the number of jets well, but has an important defect in that it produces retrograde (easterly) jets at the equator. The reason for this defect is that turbulent horizontal mixing of the planetary rotational vorticity, which has opposite signs in the two hemispheres, gives rise to a slower (retrograde) atmospheric rotation than the planetary value over the equator. The sign of the parameter β , defined as the rate of change of the Coriolis parameter with latitude, dictates the sign of the jets (see equation (2.4) in ref. 5) by the conservation of potential vorticity⁶, which measures the total strength of the atmospheric vorticity tube, in the course of this mixing process⁷.

The atmospheres of Jupiter and Saturn are much deeper than Earth's atmosphere, however, making possible completely different dynamics^{8–10} in which the flows are deep, extending from one hemisphere to another. These flows are homogeneous

in the direction of the axis of planetary rotation when the planetary interior is well mixed by thermal convection⁸. A single fluid layer, confined within the planetary sphere, therefore describes these deep flows¹⁰.

Analogous dynamics that conserve a potential vorticity govern this system, which is defined in terms of the deep fluid tubes parallel to the axis of planetary rotation. The resulting β , which now measures the rate of change in the length of fluid tubes with latitude, takes the opposite sign^{2,9}. Consequently, the equatorial jet also has an opposite sign, as has been observed. Lateral transport of stronger potential vorticity, owing to longer fluid tubes, from higher latitudes to the equator, gives rise to faster (prograde) rotation.

To demonstrate this, we run the model, assuming constant density and no convective buoyancy forcing, for a long period, starting from a random state with wind variance of 100 m s^{-1} and 200 m s^{-1} for Jupiter and Saturn, respectively. These magnitudes are comparable to those observed^{1,4}. In both cases, we obtain the prograde equatorial jets of the correct width and magnitude, and also with an indication of multiple jets (Fig. 1).

Another attractive feature of our 'deep' model is that the magnitude of β values is greater, which relaxes the wind-shear stability criterion⁹ and opens the way for obtaining steeper high-latitude jets, in accordance with observations. This is another feature that conventional shallow models can reproduce only with difficulty. Inclusion of a convective buoyancy force⁸ would give rise to steeper jets in our deep model.

Jun-Ichi Yano*, Olivier Talagrand†, Pierre Drossart‡

*Laboratoire de Météorologie Dynamique, École Polytechnique, F-91128 Palaiseau, France
e-mail: yano@cnrm.meteo.fr

†LMD, Ecole Normale Supérieure, F-75231 Paris, France

‡LESIA (CNRS-FRE2461), Observatoire de Paris, F-92125 Meudon, France

1. Ingersoll, A. P. *Science* **248**, 308–315 (1990).
2. Yano, J.-I. *Chaos* **4**, 287–297 (1994).
3. Ingersoll, A. P., Gierasch, P. J., Banfield, D., Vasavada, A. R. & Galileo Imaging Team *Nature* **403**, 630–632 (2000).
4. Cho, J. Y.-K. & Polvani, L. M. *Science* **273**, 335–337 (1996).
5. Yano, J.-I. & Flierl, G. R. *Dyn. Atmos. Ocean* **16**, 439–472 (1992).
6. Pedlosky, J. *Geophysical Fluid Dynamics* 2nd edn (Springer, New York, 1987).
7. Rhines, P. B. & Young, W. R. *J. Fluid Mech.* **122**, 347–367 (1982).
8. Busse, F. H. *Icarus* **29**, 255–260 (1976).
9. Ingersoll, A. P. & Pollard, D. *Icarus* **52**, 62–80 (1982).
10. Yano, J.-I. & Flierl, G. R. *Ann. Geophysicae* **12**, 1–18 (1994).

Competing financial interests: declared none.

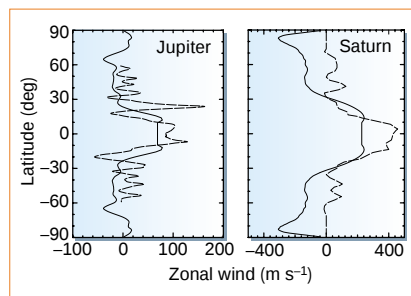


Figure 1 Simulated zonal winds (solid lines) and actual observations (dashed lines) for Jupiter and Saturn, after 300 and 100 planetary rotations, respectively, from an initial random state. Here, for numerical stability, spherical geometry is replaced by a cylindrical wall parallel to the axis of planetary rotation at 99% of the planetary radius, which intersects with the planetary surface at $\pm 8.1^\circ$ latitude. The plot assumes that homogeneous winds on this cylindrical wall are directly observed at the cloud tops of the lower latitudes. Full modelling details are available from the authors.

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com) or on request from nature@nature.com.

A globally coherent fingerprint of climate change impacts across natural systems

Camille Parmesan* & Gary Yohe†

* Integrative Biology, Patterson Laboratories 141, University of Texas, Austin, Texas 78712, USA

† John E. Andrus Professor of Economics, Wesleyan University, 238 Public Affairs Center, Middletown, Connecticut 06459, USA

Causal attribution of recent biological trends to climate change is complicated because non-climatic influences dominate local, short-term biological changes. Any underlying signal from climate change is likely to be revealed by analyses that seek systematic trends across diverse species and geographic regions; however, debates within the Intergovernmental Panel on Climate Change (IPCC) reveal several definitions of a 'systematic trend'. Here, we explore these differences, apply diverse analyses to more than 1,700 species, and show that recent biological trends match climate change predictions. Global meta-analyses documented significant range shifts averaging 6.1 km per decade towards the poles (or metres per decade upward), and significant mean advancement of spring events by 2.3 days per decade. We define a diagnostic fingerprint of temporal and spatial 'sign-switching' responses uniquely predicted by twentieth century climate trends. Among appropriate long-term/large-scale/multi-species data sets, this diagnostic fingerprint was found for 279 species. This suite of analyses generates 'very high confidence' (as laid down by the IPCC) that climate change is already affecting living systems.

The Intergovernmental Panel on Climate Change¹ (IPCC) assessed the extent to which recent observed changes in natural biological systems have been caused by climate change. This was a difficult task despite documented statistical correlations between changes in climate and biological changes^{2–5}. With hindsight, the difficulties encountered by the IPCC can be attributed to the differences in approach between biologists and other disciplines, particularly economists. Studies in this area are, of necessity, correlational rather than experimental, and as a result, assignment of causation is inferential. This inference often comes from experimental studies of the effects of temperature and precipitation on the target species or on a related species with similar habitats. Confidence in this inferential process is subjective, and differs among disciplines, thus resulting in the first divergence of opinion within the IPCC.

The second impasse came from differences in perspective on what constitutes an 'important' factor. Anyone would consider a currently strong driver to be important, but biologists also attach importance to forces that are currently weak but are likely to persist. In contrast, economic approaches tend to discount events that will occur in the future, assigning little weight to weak but persistent forces. Differences of opinion among disciplines can therefore stem naturally from whether the principal motivation is to assess the magnitude of immediate impacts or of long-term trajectories. Most field biologists are convinced that they are already seeing important biological impacts of climate change^{1–4,6–9}; however, they have encountered difficulty in convincing other academic disciplines, policy-makers and the general public. Here, we seek to improve communication, provide common ground for discussion, and give a comprehensive summary of the evidence.

How should a 'climate fingerprint' be defined? A straightforward view typical of an economist would be to conclude that climate change was important if it were principally responsible for a high proportion of current biotic changes. By this criterion a climate fingerprint appears weak. Most short-term local changes are not caused by climate change but by land-use change and by natural fluctuations in the abundance and distribution of species. This fact has been used by non-biologists to argue that climate change is of little importance to wild systems¹⁰. This approach, however, effectively ignores small, systematic trends that may become important in the longer term. Such underlying trends would be confounded (and often swamped) by strong forces such as habitat loss. Biologists

have tended to concentrate on studies that minimize confounding factors, searching for trends in relatively undisturbed systems and then testing for significant associations with climate change. Economists have viewed this as biased (nonrandom exclusion of data) whereas biologists view this as reducing non-climatic noise. Thus, economists focus on total direct evidence and apply heavy time discounting; biologists apply a 'quality control' filter to available data, accept indirect (inferential) evidence and don't apply time discounting.

The test for a globally coherent climate fingerprint does not require that any single species show a climate change impact with 100% certitude. Rather, it seeks some defined level of confidence in a climate change signal on a global scale. Adopting the IPCC 'levels of confidence'¹¹ and applying the economists' view of a fingerprint, we would have "very high confidence" in a fingerprint if we estimated that more than 95% of observed changes were principally caused by climate change, "high confidence" between 95% and 67%, "medium confidence" between 33% and 67%, and "low confidence" below 33%. In contrast, the biologists' confidence level comes from the statistical probability that global biotic trends would match climate change predictions purely by chance, coupled with supporting experimental results showing causal relationships between climate and particular biological traits.

Here, we present quantitative estimates of the global biological impacts of climate change. We search for a climate fingerprint in the overall patterns, rather than critiquing each study individually. Using the biologists' approach, we synthesize a suite of correlational studies on diverse taxa over many regions to ask whether natural systems, in general, have responded to recent climate change. Furthermore, we attempt a cross-fertilization by applying an economists' measure—the estimated proportion of observed changes for which climate trends are the principal drivers—to data sets chosen using biologists' criteria. We call this a 'global coherence' approach to the detection of climate change impacts.

First, we explore a biologists' confidence assessment with two types of analyses of observed change: statistical meta-analyses of effect size in restricted data sets and more comprehensive categorical analyses of the full literature. Second, we present a probabilistic model that considers three variables: proportion of observations matching climate change predictions, numbers of competing explanations for each of those observations, and confidence in causal

attribution of each observation to climate change. These three variables feature equally in a model that explores an economists' 'confidence' assessment. Finally, we explore diagnostic 'sign-switching' patterns that are predicted uniquely by climate change.

The evidence

A few studies indicate evolutionary responses of particular species to climate change^{12–14}, but the generality of evolutionary response remains unknown. Here, we focus on phenological (timing) shifts, range boundary shifts, and community studies on species abundances (Table 1).

Meta-analyses

We developed databases suitable for meta-analysis¹⁵ on two phenomena: range-boundary changes and phenological shifts. To control for positive publishing bias, we used only multi-species studies that reported neutral and negative results as well as positive (see Methods).

For range boundaries, suitable data spanned 99 species of birds¹⁶, butterflies¹⁷ and alpine herbs^{18,19} (see Methods). The meta-analysis showed that the range limits of species have moved on average 6.1 (± 2.4) km per decade northward or m per decade upward, significantly in the direction predicted by climate change (bootstrapped 95% confidence interval of the mean (CI_{mean}) = 1.3–10.9 km m⁻¹ per decade; one-sample *t*-test, degrees of freedom (d.f.) = 98, *t* = 2.52, *P* = 0.013; Table 2).

For phenologies, suitable data were reported for herbs^{20–23}, shrubs^{20–25}, trees^{20,23–25}, birds^{20,21}, butterflies²⁶ and amphibians^{27,28}, a total of 172 species (see Methods). There was a mean shift towards earlier spring timing of 2.3 days per decade, with a bootstrapped 95% CI of 1.7–3.2 days advancement per decade (significant at *P* < 0.05).

Categorical analyses

The remaining studies were not included in the meta-analyses, either because they were on single species or because they did not present data in the raw form of *x* unit change per *y* time units per species. These less-detailed data were simplified into four categories: changed in accord with or opposite to climate change predictions, changed in some other fashion or stable (see Methods).

As with previous studies¹⁷, analyses ignore species classified as 'stable'. This category does not represent a single result, as apparent stability could arise from a diversity of situations¹⁷ such as: 1) the phenology, abundance or distribution of the species is not driven by climatic factors; 2) the species is actually changing, but poor data resolution could not detect small changes; and 3) the phenology, abundance or distribution of the species is driven by climatic factors, but fails to respond to current climate change. Such failure could stem from anthropogenic barriers to dispersal (habitat fragmentation) or from a lag in response time. Lags are expected when limited dispersal capabilities retard poleward/upward colonization²⁹, or when a necessary resource has slower response time than the focal species¹⁷.

Phenological shifts. We quantitatively assessed 677 species reported in the literature (Table 1). Over a time period range of 16–132 years (median 45 yrs), 27% showed no trends in phenologies, 9% showed trends towards delayed spring events, whereas the remaining 62% showed trends towards spring advancement. Observed trends include earlier frog breeding^{27,28}, bird nesting^{30–32}, first flowering^{20–25}, tree budburst^{23–25}, and arrival of migrant birds and butterflies^{20,21,26,33} (Table 1). Shifts in phenologies that have occurred are overwhelmingly (87%) in the direction expected from climate change (*P* < 0.1 × 10⁻¹²; Table 2).

Distribution/abundance shifts. In a quantitative assessment covering >1,046 species, we were able to categorize 893 species, functional

Table 1 Summary of data studying phenological and distributional changes of wild species

Taxon	Ref. number	Total no. of species (or species groups)	Spatial scale			Time scale (range years)	Change in direction predicted (<i>n</i>)	Change opposite to prediction (<i>n</i>)	Stable (<i>n</i>)	No prediction (<i>n</i>)
			L	R	C					
Phenological changes										
Woody plants	20,23,24*,25*	<i>n</i> = 38 sp	2	1		35–132	30	1	7	–
Herbaceous plants	20,21*	<i>n</i> = 38 sp	1	1		63–132	12	–	26	–
Mixed plants	22*	<i>n</i> = 385 sp	1			46	279	46	60	–
Birds	20,21*,30,31,32,33	<i>n</i> = 168 sp	2	3	1	21–132	78	14	76	–
Insects	26	<i>n</i> = 35 sp		1		23	13	–	22	–
Amphibians	27,28	<i>n</i> = 12 sp	2			16–99	9	–	3	–
Fish	20	<i>n</i> = 2 sp	1			132	2	–	–	–
Distribution/abundance changes										
Tree lines	54,55,56*	<i>n</i> = 4 sp + 5 grps	2	1		70–1,000	3 sp + 5 grps	–	1	–
Herbs and shrubs	18,19,41*,42*	<i>n</i> > 66 sp, 15 detailed		3		28–80	13	2	–	–
Lichens	36	4 biogeographic grps (<i>n</i> = 329 sp)	1			22	43	9	113	164
Birds	8*	<i>n</i> = 3 sp	1			50	3	–	–	–
	16,57*	N sp (<i>n</i> = 46 sp)	2			20–36	13	15	18	–
		S sp (<i>n</i> = 73 sp)	2			20–36	36	16	21	6
	43*	Low elevation (>91 sp)	1			20	71	11	9	–
		High elevation (>96 sp)	1			20	37	27	32	–
Mammals	37	<i>n</i> = 2 sp		1		52	2	–	–	–
Insects	17,49*	<i>n</i> = 36 sp	1	1		98–137	23	2	10	1
	17	N boundaries (<i>n</i> = 52 sp)	1			98	34	1	17	–
		S boundaries (<i>n</i> = 40 sp)	1			98	10	2	28	–
Reptiles and amphibians	43*	<i>n</i> = 7 sp	1			17	6	–	1	–
Fish	39	4 biogeographic grps (<i>n</i> = 83 sp)	1			–	2 grps	–	1 grp	1 grp
	40*	N sp (<i>n</i> > 1 sp)	1			70	>1	–	–	–
		S sp (<i>n</i> > 1 sp)	1			70	>1	–	–	–
Marine invertebrates	34*,40*	N sp (<i>n</i> > 21)	1	1		66–70	>19	2	–	>1 sp not classified
		S sp (<i>n</i> > 21)	1	1		66–70	>20	1	–	–
		Cosmopolitan sp (<i>n</i> = 28 sp)	1			66	–	–	–	28
Marine zooplankton	40*	Cold water (<i>n</i> > 10 sp)	1			70	>10	–	–	>8 sp not classified
		Warm water (<i>n</i> > 14 sp)	1			70	>14	–	–	–
	35	6 biogeographic grps (<i>n</i> ≥ 36 sp)		1		39	6 grps	–	–	–

N, species with generally northerly distributions (boreal/arctic); S, species with generally southerly distributions (temperate); L, local; R, regional (a substantial part of a species distribution; usually along a single range edge); C, continental (most or the whole of a species distribution). No prediction indicates that a change may have been detected, but the change was orthogonal to global warming predictions, was confounded by non-climatic factors, or there is insufficient theoretical basis for predicting how species or system would change with climate change.

* Study partially controlled for non-climatic human influences (for example, land-use change). Studies that were highly confounded with non-climatic factors were excluded. (See Supplementary Information for details of species classification.)

groups or biogeographic groups (Table 1). Less than one-third (27%) of these have exhibited stable distributions during the twentieth century. Others (24%) show changes that are impossible to relate to climate change predictions. These two types of result neither support nor refute a climate change signal, although it will be important for predictive biological models to eventually determine what proportion of these are truly stable systems.

Some range shifts have been measured directly at range boundaries, whereas others have been inferred from abundance changes within local communities. Over all of the range and abundance shift data, 434 species were categorized as changing over time periods of 17–1,000 years (median 66 years) (Table 1). Of these, 80% have shifted in accord with climate change predictions (see Methods) ($P < 0.1 \times 10^{-12}$; Table 2). New species have colonized previously 'cool' regions, including sea anemones in Monterey Bay³⁴ and lichens and butterflies in Europe^{17,36}, whereas some Arctic species have contracted in range size^{35,37}. Over the past 40 years, maximum range shifts vary from 200 km (butterflies¹⁷) to 1,000 km (marine copepods³⁴).

Probabilistic coherence

How strong is the climate change signal in the light of confounding factors and lack of experimentation? We investigate this argument in a probabilistic context. We formulated a probabilistic model to ask whether a climate change fingerprint exists in a disparate set of n observed biological changes. Let n'/n indicate the proportion of observations counter to climate change predictions and p indicate the probability that climate change is the only possible causal agent of the observed biological change in any of the $n - n'$ species that do conform to climate change predictions. In practice, this can be estimated across a set of species by assigning each species a 0 or a 1, depending on whether or not competing explanations exist; p then is the proportion of species that have no competing explanations.

Competing (non-climatic) explanations can, therefore, be expected in $\{(1 - p)(n - n')\}$ of the reported analyses. Finally, for any of the $n - n'$ climate-conforming species, let π indicate the probability, determined from previous empirical study, that climate change is the principal causal agent of a particular biological change (independent of p).

These three variables, each varying from 0 to 1, are inputs to a binomial probability model whose output estimates the proportion of all species that are, in truth, being impacted by climate change. In practice, confounding factors can never be eliminated completely from observational studies; therefore, p would normally have a low value. Here, we consider only the conservative case where $p = 0$; that is, we assume that non-climatic alternative explanations exist for every species. In the Supplementary Information, we present modelling schemes where p varies from 0 to 1.0.

The importance of non-climatic explanations should decrease

with increasing scale. Most local changes are idiosyncratic and consist of noise when scaled up; however, atmospheric carbon dioxide levels have risen nearly uniformly across the globe. Increased CO₂ can directly cause earlier flowering³⁸, as does increased temperature, making these effects difficult to separate. However, these two effects can be viewed as different aspects of global warming, legitimizing discussion of their joint impacts.

The variable π reflects the extent to which previous study and experimentation provides clear mechanistic understanding of the links between climate variables and a species' behaviour and ecology. To understand the importance of π , consider the case of the silver-spotted skipper butterfly (*Hesperia comma*) that has expanded its distribution close to its northern boundary in England over the past 20 years. Possible ecological explanations for this expansion are regional warming and changes in land use. Comparing the magnitudes and directions of these two factors suggests that climate change is more likely than land-use change to be the cause of expansion²⁹. Deeper support was provided by previous empirical studies documenting strong thermal limitation. At the northern boundary, development of offspring was restricted to the hottest microclimates (south-facing chalk slopes). Range expansion coincided with colonization of non-southern slopes. Simulation models based solely on previously measured thermal tolerances (that is, without land-use change) closely matched the observed expansion of 16.4 km (model prediction 14.4 km)¹². Thus, mechanistic understanding of the system generates a high estimate for π .

Figure 1 shows relationships between the n'/n proportions and the minimum value of π that would be required to sustain different degrees of confidence for $p = 0$. For example, the medium confidence region shows minimum values of π that would be required across the displayed range of n'/n proportions to guarantee that about half of the observed species impacts were in truth being driven principally by climate change. Claiming a climate fingerprint with high confidence would require high minimum values for π (>0.67) regardless of n'/n .

Applying the probabilistic model

Using all of the data from Table 2 to parameterize the model, $n' = 147$ and $n = 770$, making $n'/n = 0.16$ (16% of species changing opposite to climate change predictions). We now consider π . The extent to which climate change can be isolated as the predominant driving force is extremely variable among species and systems. Such attribution results from a subjective synthesis of experimental and observational research, often conducted well before and independently of any study of long-term trends. The species for which π is high are those with a history of basic biological research, especially where research has been conducted along several axes (controlled laboratory/greenhouse experiments, field manipulations and observations).

Table 2 Summary statistics and synthetic analyses derived from Table 1

Type of change	Changed as predicted	Changed opposite to prediction	P-value
Phenological ($N = 484/(678)$)	87% ($n = 423$)	13% ($n = 61$)	$<0.1 \times 10^{-12}$
Distributional changes			
At poleward/upper range boundaries	81%	19%	—
At equatorial/lower range boundaries	75%	25%	—
Community (abundance) changes			
Cold-adapted species	74%	26%	—
Warm-adapted species	91%	9%	—
$N = 460/(920)$	81% ($n = 372$)	19% ($n = 88$)	$<0.1 \times 10^{-12}$
Meta-analyses			
Range-boundaries ($N = 99$)	6.1 km m ⁻¹ per decade northward/upward shift*		0.013
Phenologies ($N = 172$)	2.3 days per decade advancement*		<0.05

Data points represent species, functional groups or biogeographic groups. N , number of statistically or biologically significant changes/(total number species with data reported for boundary, timing, or abundance processes). The no prediction category is not included here.

* Bootstrap 95% confidence limits for mean range boundary change are 1.26, 10.87; for mean phenological shift the limits are -1.74, -3.23.

This sort of biological detail reveals that climate and extreme weather events are mechanistically linked to body size, individual fitness and population dynamics for diverse species^{3–9} (but not for all). Species for which confidence in climate as the primary driving mechanism is low are those for which long-term observational records exist, but not detailed empirical research on target species or on ecologically similar species. The black line in Fig. 1 suggests that medium confidence can be claimed for $n'/n = 0.16$ if $0.35 < \pi < 0.7$. Other contingencies, such as complications from a positive publishing bias or non-independence among confounding factors, can be considered through variations of the model (see Supplementary Information).

Differentiating diagnostic patterns

Predictions of the impacts of climate change are not unidirectional, but may show opposite trends within communities and across long time spans or large spatial scales. Alternative causal agents would therefore have to be able to switch the sign of their impacts within a study if they were to form credible competing explanations. Such differentiating patterns greatly reduce the likelihood of hidden, non-climate competing explanations, thereby increasing P and decreasing the value of π necessary to achieve a given confidence level (see Supplementary Information). High confidence could be obtained under this scheme with existing patterns ($n'/n \leq 0.33$) and poor mechanistic understanding (low π). Sufficient data to quantify the differential impacts on species' distributions or phenologies across time periods or geographic regions were available for 334 species, among which 84% showed a sign-switching diagnostic of climate change response ($P < 0.1 \times 10^{-12}$; Table 3).

Community representation sign switching

Community studies in regions of overlapping 'polar' and 'temperate' species base their climate change attribution on differential responses of these two categories. Among marine fish and intertidal invertebrates (for example, snails, barnacles, anemones, copepods and limpets) off the Californian coast^{34,39} and in the North Atlantic^{35,40}, lichens in the Netherlands³⁶, foxes in Canada³⁷ and birds in Great Britain¹⁶, polar species have tended to be stable or decline in abundance, whereas temperate species at the same site have increased in abundance and/or expanded their distributions. Analogous shifts are occurring even within the Arctic and Antarctic among penguins⁸, woody plants⁴¹ and vascular plants⁴². Similar patterns

exist for lowland compared with highland birds in the tropics⁴³. Most of these studies are local, with high variability of individual species' population dynamics. Even so, 80% of changes in community representation are in accord with climate change predictions (Tables 2 and 3).

Temporal sign switching

Long-term studies encompass periods of climate cooling as well as warming. If the distributions of species are truly driven by climate trends, these species should show opposite responses to cooling and warming periods. Such sign switching has been documented in the United Kingdom for marine fish, limpets, barnacles and zooplankton⁴⁰, in the United Kingdom and Estonia for birds^{20,31,44,45}, and in the United Kingdom, Finland and Sweden for butterflies^{17,46–48} (see also Table 3 legend). A typical pattern includes northward range shifts during the two twentieth-century warming periods (1930–45 and 1975–99), and southward shifts during the intervening cooling period (1950–70). No species showed opposing temporal trends (Table 3).

Spatial sign switching

Whole-range, continental-scale studies, by encompassing the extremes of a species' distribution, allow testing for differential spatial impacts. In North America and Europe, detailed temporal data spanning the twentieth century were compiled for 36 butterfly species at both northern and southern range extremes^{17,49}. Eight species (22%) exhibited a diagnostic pattern of northward expansion (new colonizations) and southern contraction (population extinctions). No species showed opposing range shift trends (northward contraction and southward expansion) (Table 3).

Discussion

The logic of a global focus on biological change is analogous to that for climate change itself. With climate change, attribution of recent warming trends to changes in atmospheric gases comes from analysis of global patterns, not from detailed data from individual meteorological stations. Similarly, when assessing biological

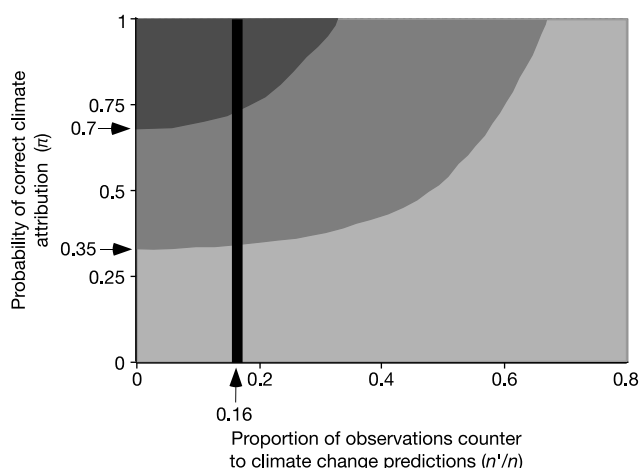


Figure 1 Probabilistic model based on parameter estimates from a review of the literature. Levels of confidence in the linkage of biological changes to global climate change are: high (dark grey), medium (mid-grey) and low (light grey). Confidence regions assume $p = 0$ (competing explanations exist for all studies). The black line indicates the region of confidence possible using the probabilistic model on the basis of the parameter estimate of n'/n from the literature review, and allowing π to vary freely.

Table 3 Biological fingerprint of climate change impacts

Sign-switching pattern	Percentage of species showing diagnostic pattern
Community	
Abundance changes have gone in opposite directions for cold-adapted compared with warm-adapted species. Usually local, but many species in each category. Diverse taxa, $n = 282^*$.	80%
Temporal	
Advancement of timing of northward expansion in warm decades (1930s/40s and 1980s/90s); delay of timing or southward contraction in cool decades (1950s/60s), 30–132 years per species. Diverse taxa, $n = 44^*$.	100%
Spatial	
Species exhibit different responses at extremes of range boundary during a particular climate phase. Data are from substantial parts of both northern and southern range boundaries for each species. All species are northern hemisphere butterflies, $n = 8^*$.	100%

Differential sign-switching patterns diagnostic of climate change as the underlying driver.
^{*}Numbers of species represent minimum estimates, as not all species were described in sufficient detail in each study to classify. A few species showed two types of sign switching, and so are included in more than one cell. Data are from references in text and from raw data provided by L. Kaila, J. Kullberg, J. J. Lennon, N. Ryrholm, C. D. Thomas, J. A. Thomas and M. Warren.

impacts, the global pattern of change is far more important than any individual study.

The approach of biologists selects study systems to minimize confounding factors and deduces a strong climate signal both from systematic trends across studies and from empirically derived links between climate and biological systems. This deduction is made even if climate explains only a small part of the observed biological change. The meta-analyses of 334 species and the global analyses of 1,570 species (or functional/biogeographic groups) show highly significant, nonrandom patterns of change in accord with observed climate warming in the twentieth century, indicating a very high confidence (>95%) in a global climate change fingerprint (Table 2).

The approach of economists takes a broader view. In its purest form, applied to all existing data and incorporating time discounting, this approach would conclude that climate change has little total impact on wild species. We argue that this approach misses biologically important phenomena. Here we hybridize the two approaches by applying an economists' model to data that biologists would consider reasonable, and forego time discounting. A total of 74–91% of species that have changed have done so in accord with climate change predictions (Table 2) giving an estimate of $n'/n = 0.16$ for the hybrid model. Assessment of π , the probability of correct attribution to climate, is subjective and relies on the level of confidence in inferential evidence. Such evidence comes from empirical analyses and experimental manipulations, which have documented the importance of climatic variables to the dynamics, distributions and behaviour of species^{3,5,8,9}. From these studies, biologists infer that expected values of π are often high. We show that moderate values of π (0.35–0.70) are consistent with medium confidence in a global climate change fingerprint.

The different approaches raise two distinct questions of the data and result in different levels of confidence in a climate change fingerprint. The questions are: (1) whether climate change can be shown to be an over-riding factor currently driving natural systems; and (2) whether there is sufficient evidence to implicate climate change as a common force impacting natural systems on a global scale. In an absolute sense, land-use change has probably been a stronger driver of twentieth century changes in wild plants and animals than has climate change (question 1). From a biological view, however, finding any significant climate signal amidst noisy biological data is unexpected in the absence of real climate drivers (question 2). Such small, persistent forces are inherently important in that they can alter species interactions, de-stabilize communities and drive major biome shifts.

A review of the literature reveals that the patterns that are being documented in natural systems are surprisingly simple, despite the real and potential complexity of biotic change. Change in any individual species, taxon or geographic region may have a number of possible explanations, but the overall effects of most confounding factors decline with increasing numbers of species/systems studied. Similarly, uncertainty in climate attribution for any particular study does not prevent the development of a global conclusion on the basis of a cumulative synthesis. In particular, a clear pattern emerges of temporal and spatial sign switches in biotic trends uniquely predicted as responses to climate change. With 279 species (84%) showing predicted sign switches, this diagnostic indicator increases confidence in a climate change fingerprint from either viewpoint.

The published IPCC conclusion stated high confidence ($P > 0.67$) in a climate signal across observed biotic and abiotic changes. Analyses presented here support that conclusion. Furthermore, a driver of small magnitude but consistent impact is important in that it systematically affects century-scale biological trajectories and ultimately the persistence of species. The climate fingerprint found here implicates climate change as an important driving force on natural systems. □

Methods

Climate change predictions

Expected phenological shifts for regions experiencing warming trends are for earlier spring events (for example, migrant arrival times, peak flight date, budburst, nesting, egg-laying, and flowering) and for later autumn events (for example, leaf fall, migrant departure times, and hibernation)^{50,51}. Response to climate warming predicts a preponderance of poleward/upward shifts^{50,51}. Dynamics at the range boundaries are expected to be more influenced by climate than are dynamics within the interior of a species range. Thus, community level studies of abundance changes are used best to infer range shifts when they are located at ecotones involving species having fundamentally different geographic ranges: higher compared with lower latitudes, or upper compared with lower altitudes. Response to climate warming predicts that southerly species should outperform northerly species at the same site^{50,51}.

Selection of studies for review

This was not an exhaustive review. The studies listed in Table 1 comprise the bulk of wild species studied with respect to climate change hypotheses. Selection of papers was aimed at those with one or more of the following attributes: long temporal span (>20 years), data covering a large geographic region, and/or data gathered in an unbiased manner for a multi-species assemblage (typically species abundance data of locally well-documented communities). We excluded several high-quality studies of single species performed at local scale or highly confounded by non-climatic global change factors. The stable category represents species for which any observed changes are indistinguishable from year to year fluctuations, either from a statistical test for trend using very long time series data or from comparing net long-term movement to expected yearly variation on the basis of basic biological knowledge of dispersal/colonization abilities.

Meta-analyses

To create databases, studies were combined that made similar types of measurements and that reported quantitative estimates of change over a specified time period. All species were used; that is, even species that are categorized as stable in Table 1 were included in the meta-analysis. We treated phenological and distributional changes separately. To minimize positive publishing bias, only multi-species studies were included.

We considered each species as an independent data point, rather than each study. Only data reported in terms of change per individual species were included. This precluded use of studies that only report mean change across a set of species.

We used only distributional studies at range boundaries. We excluded equatorial and lower elevational boundaries because of a paucity of data combined with theoretical reasons for treating these boundaries separately from poleward/upper elevational boundaries⁵². Three studies met the criteria for data detail, covering 9 alpine herbs^{18,19}, 59 birds¹⁶ and 31 butterflies¹⁷. The geographic locations of these boundaries were non-overlapping, reducing the likelihood of correlated confounding variables. Altitude was converted to latitudinal equivalent (for temperature clines, 1 km northward = 1 m upward). The United Kingdom bird data compared mean northern boundary in 1999 to that in 1972 using the ten northernmost occupied grid cells (on 10 km² grids) from published atlases. The Swedish butterfly data compared mean northern boundary in the period 1971–97 to mean northern boundary in 1900–20 using the five northernmost records per year. The Swiss herb data showed changes in species assemblages over the twentieth century in fixed plots up altitudinal gradients on 26 mountains.

The effect size per species was the absolute magnitude of range boundary shift, standardized across species to be in units of km m⁻¹ per decade, with northward/upslope shifts positive and southward/downslope shifts negative. Data were not skewed, and n was large. Therefore, a one-sample t -test was used to evaluate the null hypothesis of no overall trends (that is, H_0 : mean boundary change across all species is zero). Variances were not available for all species, so we used an unweighted analysis. We performed an additional bootstrap analysis of 95% confidence limits on the mean boundary shift (10,000 iterations)⁵³.

The phenological meta-analysis was on spring timing events—there were insufficient studies on autumn phenology to warrant analysis. Nine studies published magnitudes of shift over a given time period (17–61 years). They included 11 trees^{20,23–25}, 6 shrubs^{20,21,23–25}, 85 herbs^{20–23}, 35 butterflies²⁶, 21 birds²¹, 12 amphibians^{27,28} and 2 fish²⁰. This data set was inappropriate for the t -test owing to skew, but bootstrapped confidence limits provided an estimate of the probability that the true mean shift includes zero.

For both analyses, geography and taxa are confounded. For the range boundary analysis, all bird data are from the United Kingdom, all butterfly data from Sweden, and all herb data from Switzerland. For the phenological analysis, most shrub and bird data are from the United States, butterfly data from Great Britain, and trees from Europe. Therefore, it is not meaningful to split the analyses further.

Categorical analyses

Reported data from all studies listed in Tables 1 and 3 were included in the categorical analyses. The predicted direction is a change predicted by global warming scenarios^{50,51}. All studies were conducted in temperate Northern Hemisphere, except for 194 species in Costa Rica⁴³ and 5 species in Antarctica⁴². Two categories showing changes either predicted by or opposite to predictions of climate change theory were tested against the random expectation of an equal probability of observing changes in either direction. Analyses were by binomial test with H_0 : $P = 0.5$.

Received 5 March; accepted 22 October 2002; doi:10.1038/nature01286.

1. Intergovernmental Panel on Climate Change Third Assessment Report *Climate Change 2001: Impacts, Adaptation, and Vulnerability* (eds McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J. & White, K. S.) (Cambridge Univ. Press, Cambridge, 2001).

2. Easterling, D. R. *et al.* Climate extremes: observations, modeling, and impacts. *Science* **289**, 2068–2074 (2000).
3. Parmesan, C., Root, T. L. & Willig, M. Impacts of extreme weather and climate on terrestrial biota. *Bull. Am. Meteorol. Soc.* **81**, 443–450 (2000).
4. Pounds, J. A. Climate and amphibian declines. *Nature* **410**, 639–640 (2001).
5. Otterson, G. *et al.* Ecological effects of the North Atlantic Oscillation. *Oecologia* **128**, 1–14 (2001).
6. Walther, G.-R. *et al.* Ecological responses to recent climate change. *Nature* **416**, 389–395 (2002).
7. Peñuelas, J. & Flella, I. Responses to a warming world. *Science* **294**, 793–795 (2001).
8. Smith, R. C. *et al.* Marine ecosystem sensitivity to climate change. *Biol. Sci.* **49**, 393–404 (1999).
9. Hoegh-Guldberg, O. Climate change, coral bleaching and the future of the world's coral reefs. *Mar. Freshwater Res.* **50**, 839–866 (1999).
10. Lomborg, B. *The Skeptical Environmentalist* (Cambridge Univ. Press, Cambridge, 2001).
11. Moss, R. & Schneider, S. *Cross Cutting Issues Guidance Papers* Intergovernmental Panel on Climate Change (World Meteorological Organization, Geneva, 2000).
12. Thomas, C. D. *et al.* Ecological processes at expanding range margins. *Nature* **411**, 577–581 (2001).
13. Rodríguez-Trellis, F. & Rodríguez, M. A. Rapid micro-evolution and loss of chromosomal diversity in *Drosophila* in response to climate warming. *Evol. Ecol.* **12**, 829–838 (1998).
14. de Jong, P. W. & Brakefield, P. M. Climate and change in clines for melanism in the two-spot ladybird, *Adalia bipunctata* (Coleoptera: Coccinellidae). *Proc. R. Soc. Lond. B* **265**, 39–43 (1998).
15. Gurevitch, J. & Hedges, L. V. *Design and Analysis of Ecological Experiments* 2nd edn (eds Scheiner, S. M. & Gurevitch, J.) 347–370 (Oxford Univ. Press, Oxford, 2001).
16. Thomas, C. D. & Lennon, J. J. Birds extend their ranges northwards. *Nature* **399**, 213 (1999).
17. Parmesan, C. *et al.* Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**, 579–583 (1999).
18. Grabherr, G., Gottfried, M. & Pauli, H. Climate effects on mountain plants. *Nature* **369**, 448 (1994).
19. Grabherr, G., Gottfried, M., Gruber, A. & Pauli, H. *Arctic and Alpine Biodiversity* (eds Chapin, F. S. III & Körner, C.) 167–181 (Springer, Berlin, 1995).
20. Ahas, R. Long-term phyto-, ornitho- and ichthyophenological time-series analyses in Estonia. *Int. J. Biometeorol.* **42**, 119–123 (1999).
21. Bradley, N. L., Leopold, A. C., Ross, J. & Huffaker, W. Phenological changes reflect climate change in Wisconsin. *Proc. Natl Acad. Sci. USA* **96**, 9701–9704 (1999).
22. Fitter, A. H. & Fitter, R. S. R. Rapid changes in flowering time in British plants. *Science* **296**, 1689–1691 (2002).
23. Menzel, A., Estrella, N. & Fabian, P. Spatial and temporal variability of the phenological seasons in Germany from 1951 to 1996. *Glob. Change Biol.* **7**, 657–666 (2001).
24. Menzel, A. & Fabian, P. Growing season extended in Europe. *Nature* **397**, 659 (1999).
25. Menzel, A. Trends in phenological phases in Europe between 1951 and 1996. *Intl J. Biometeorol.* **44**, 76–81 (2000).
26. Roy, D. B. & Sparks, T. H. Phenology of British butterflies and climate change. *Glob. Change Biol.* **6**, 407–416 (2000).
27. Beebe, T. J. C. Amphibian breeding and climate. *Nature* **374**, 219–220 (1995).
28. Gibbs, J. P. & Breisch, A. R. Climate warming and calling phenology of frogs near Ithaca, New York, 1900–1999. *Conserv. Biol.* **15**, 1175–1178 (2001).
29. Warren, M. S. *et al.* Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature* **414**, 65–69 (2001).
30. Crick, H. Q. P., Dudley, C., Glue, D. E. & Thomson, D. L. UK birds are laying eggs earlier. *Nature* **388**, 526 (1997).
31. Crick, H. Q. P. & Sparks, T. H. Climate related to egg-laying trends. *Nature* **399**, 423–424 (1999).
32. Dunn, P. O. & Winkler, D. W. Climate change has affected the breeding date of tree swallows throughout North America. *Proc. R. Soc. Lond. B* **266**, 2487–2490 (1999).
33. Gatter, W. Zugzeiten und Zugmuster im Herbst: Einfluss des Treibhauseffekts auf den Vogelzug? *J. Ornithol.* **133**, 427–436 (1992).
34. Sagarin, R., Barry, J. P., Gilman, S. E. & Baxter, C. H. Climate-related change in an intertidal community over short and long time scales. *Ecol. Monogr.* **69**, 465–490 (1999).
35. Beaupré, G., Reid, P. C., Ibañez, F., Lindley, J. A. & Edwards, M. Reorganization of North Atlantic marine copepod biodiversity and climate. *Science* **296**, 1692–1694 (2002).
36. van Hark, C. M., Aptroot, A. & van Dobben, H. F. Long-term monitoring in the Netherlands suggests that lichens respond to global warming. *Lichenologist* **34**, 141–154 (2002).
37. Hersteinsson, P. & MacDonald, D. W. Interspecific competition and the geographical distribution of

- Red and Arctic foxes *Vulpes vulpes* and *Alopex lagopus*. *Oikos* **64**, 505–515 (1992).
38. Rusterholz, H. P. & Erhardt, A. Effects of elevated CO₂ on flowering phenology and nectar production of nectar plants important for butterflies of calcareous grasslands. *Oecologia* **113**, 341–349 (1998).
39. Holbrook, S. J., Schmitt, R. J. & Stephens, J. S. Changes in an assemblage of temperate reef fishes associated with a climate shift. *Ecol. Appl.* **7**, 1299–1310 (1997).
40. Southward, A. J., Hawkins, S. J. & Burrows, M. T. Seventy years' observations of changes in distribution and abundance of zooplankton and intertidal organisms in the western English Channel in relation to rising sea temperature. *J. Thermal Biol.* **20**, 127–155 (1995).
41. Sturm, M., Racine, C. & Tape, K. Increasing shrub abundance in the Arctic. *Nature* **411**, 546–547 (2001).
42. Smith, R. I. L. Vascular plants as bioindicators of regional warming in Antarctica. *Oecologia* **99**, 322–328 (1994).
43. Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. Biological responses to climate change on a tropical mountain. *Nature* **398**, 611–615 (1999).
44. Williamson, K. Birds and climatic change. *Bird Study* **22**, 143–164 (1975).
45. McCleery, R. H. & Perrins, C. M. Temperature and egg-laying trends. *Nature* **391**, 30–31 (1998).
46. Asher, J., *et al.* *The Millennium Atlas of Butterflies in Britain and Ireland* (Oxford Univ. Press, Oxford, 2001).
47. Dennis, R. L. H. *Butterflies and Climate Change* (Manchester Univ. Press, Manchester, 1993).
48. Henriksen, H. J. & Kretzsch, I. B. *The Butterflies of Scandinavia in Nature* (Skandinavisk Bogforlag, Denmark, 1982).
49. Parmesan, C. Climate and species range. *Nature* **382**, 765–766 (1996).
50. Peters, R. L. *Global Warming and Biological Diversity* (eds Peters, R. L. & Lovejoy, T. E.) (Yale Univ. Press, New Haven, 1992).
51. Schneider, S. H. *Biotic Interactions and Global Change* (eds Kareiva, P. M., Kingsolver, J. G. & Huey, R. B.) (Sinauer, Sunderland, Massachusetts, 1993).
52. MacArthur, R. H. *Geographical Ecology* (Harper and Row, New York, 1972).
53. Aksenov, S. V. Mathematica package for confidence intervals by Bootstrap v.1.12 (Wolfram Research, Mathematica version 4, Champaign, Illinois, 2002).
54. Kullman, L. 20th century climate warming and tree-limit rise in the southern Scandes of Sweden. *AMBIO* **30**, 72–80 (2001).
55. Payette, S., Filion, L., Delwaide, A. & Bégin, C. Reconstruction of tree-line vegetation response to long-term climate change. *Nature* **341**, 429–432 (1989).
56. Ross, M. S., O'Brien, J. J., Da Silveira, L. & Lobo Sternberg, L. Sea-level rise and the reduction in pine forests in the Florida Keys. *Ecol. Appl.* **4**, 144–156 (1994).
57. Johnson, J. R. Jr *A Century of Avifaunal Change in Western North America* (eds Jehl, J. R. & Johnson, N. K.) (Cooper Ornithological Society, Lawrence, Kansas, 1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This paper was stimulated by discussion during meetings of the Intergovernmental Panel on Climate Change, particularly with Q. K. Ahmad, N. Leary, R. Leemans, R. Moss, J. Price, T. L. Root, C. Rosenzweig, S. Schneider, R. Tol, F. Toth and R. Warrick. We thank L. Kaila, J. Kullberg, J. J. Lennon, N. Ryrholm, C. D. Thomas, J. A. Thomas and M. Warren for use of their raw data for analyses. We also thank C. Krebs, J. Matthews, R. Plowes, J. A. Pounds, R. Sagarin, M. C. Singer and B. Wee. Writing was facilitated by the Centre National de la Recherche Scientifique (CEFE) and by the National Science Foundation of the United States through its support of the Center for Integrated Assessment of the Human Dimensions of Global Change at Carnegie Mellon University.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.P. (e-mail: parmesan@mail.utexas.edu).

Wave-like properties of solar supergranulation

L. Gizon*, T. L. Duvall Jr† & J. Schou*

* W. W. Hansen Experimental Physics Laboratory, Stanford University, Stanford, California 94305, USA

† Laboratory for Astronomy and Solar Physics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Supergranulation^{1,2} on the surface of the Sun is a pattern of horizontal outflows, outlined by a network of small magnetic features, with a distinct scale of 30 million metres and an apparent lifetime of one day. It is generally believed that supergranulation corresponds to a preferred ‘cellular’ scale of thermal convection; rising magnetic fields are dragged by the outflows and concentrated into ‘ropes’ at the ‘cell’ boundaries³. But as the convection zone is highly turbulent and stratified, numerical modelling has proved to be difficult and the dynamics remain poorly understood. Moreover, there is as yet no explanation for the observation that the pattern appears^{4,5} to rotate faster around the Sun than the magnetic features. Here we report observations showing that supergranulation undergoes oscillations and supports waves with periods of 6–9 days. The waves are predominantly prograde, which explains the apparent super-rotation of the pattern. The rotation of the plasma through which the pattern propagates is consistent with the motion of the magnetic network.

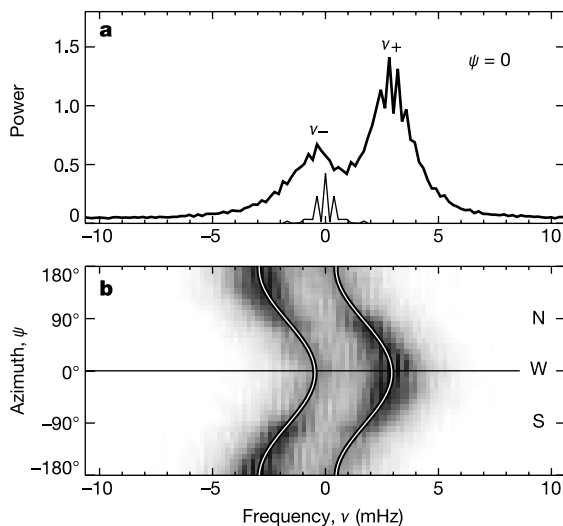


Figure 1 Power spectrum of the supergranulation signal near the solar equator ($\lambda = 0^\circ$). Sections are shown at constant wavenumber $k = 120/R$, where R is the solar radius. **a**, The thick line is the power spectrum versus frequency, ν , for $\mathbf{k} = (k, 0)$ pointing in the direction of solar rotation, $\psi = 0$ (that is, west). There are two peaks, at frequencies ν_- and ν_+ . The frequency resolution is given by the power spectrum of the temporal window function (thin line). **b**, Cylindrical section, $P_k(\nu, \psi)$, of the power spectrum at constant k , plotted against ν and the direction of $\mathbf{k}$, ψ . By construction, $P_k(\nu, \psi) = P_k(-\nu, \psi - \pi)$. Power peaks in two ridges at frequencies $\nu_-(\psi)$ and $\nu_+(\psi)$. For each ψ , we measure $\nu_\pm$ by fitting the sum of two independent lorentzian functions to the power. The fits take into account the convolution by the window function. The sinusoidal variation of $\nu_\pm$ with ψ is due to advection by a background flow $\mathbf{u} = (u_x, u_y)$. The double lines show the fit $\nu = \pm \nu_0 + (ku_x \cos \psi + ku_y \sin \psi)/2\pi$ to $\nu_\pm(\psi)$, where ν_0 is a constant frequency. At the equator we find $\mathbf{u} = (43, 0) \text{ m s}^{-1}$. The velocity u_x is measured in a frame co-rotating with the Sun at the Carrington rotation rate.

To study supergranulation, we use a 60-day sequence of Doppler velocity images obtained in 1996 by the Michelson Doppler Imager⁶ (MDI) on board the Solar and Heliospheric Observatory (SOHO). Images are tracked at the Carrington angular velocity ($\Omega_C = 2.87 \mu\text{rads}^{-1}$) to remove the main component of solar rotation. We apply the techniques of time-distance helioseismology⁷ to obtain every 12 h a $120^\circ \times 120^\circ$ map of the horizontal divergence of the flows in a 1-Mm-deep layer beneath the surface⁸. Unlike raw Doppler images, the divergence signal has uniform sensitivity across the solar disk and is subject to few systematic errors. Supergranules appear as cellular patterns of horizontal outward flow in the divergence maps.

For any given target latitude, λ , we extract a longitudinal section of the data 10° -wide in latitude centred about λ . The divergence signal is Fourier-transformed in three dimensions to make power spectra as a function of frequency, ν , and horizontal wavevector, $\mathbf{k} = (k_x, k_y)$, where k_x and k_y are in the east–west and south–north directions, respectively. In cylindrical coordinates, $\mathbf{k}$ is uniquely specified by its magnitude, k , and its direction, ψ , such that $k_x = k \cos \psi$ and $k_y = k \sin \psi$. Fig. 1 shows sections of the equatorial power spectrum at a constant k typical of the supergranulation. For each azimuth ψ , the power has two broad peaks at frequencies ν_+ and ν_- (Fig. 1a). No galilean transformation can cause these peaks to coalesce, at zero frequency or otherwise. This implies that the supergranulation undergoes oscillations.

Observations show that the frequencies $\nu_\pm$ have a sinusoidal dependence, with ψ of the form $\pm \nu_0 + \nu_1 \cos(\psi - \psi_0)$ (Fig. 1b). We interpret ν_1 to be a Doppler frequency shift, $\nu_1 = k\|\mathbf{u}\|/2\pi$, produced by a horizontal background flow $\mathbf{u}$ with magnitude $\|\mathbf{u}\|$ and pointing in the direction ψ_0 , as in helioseismological ring analysis⁹. The nearly linear relationship measured between ν_1 and k in the range $40 < kR < 180$, where R is the solar radius, is consistent with this interpretation. The latitudinal dependence of $\mathbf{u}$ is shown in Fig. 2. The inferred rotation (Fig. 2a) and meridional circulation (Fig. 2b) are both remarkably similar to that of the small magnetic features^{10,11}. This property is consistent with the view that magnetic fields are advected by supergranular flows.

The dynamics of the supergranulation is best studied once the background flow, $\mathbf{u}$, has been removed. In a co-moving frame, each spatial component oscillates at a characteristic frequency ν_0 . We find a clear relationship between ν_0 and the wavenumber k , well described by a power law (Fig. 3a). This is a fundamental relationship, as it is measured to be independent of both ψ and λ . The data are consistent with a spectrum of travelling waves with a dispersion

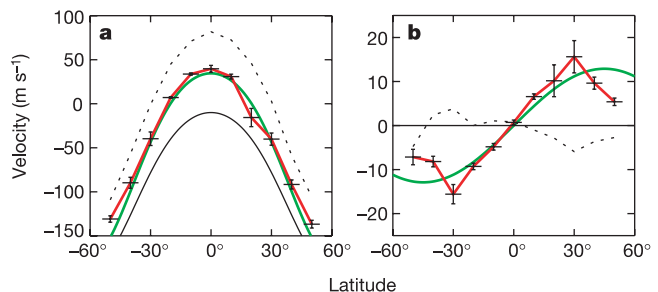


Figure 2 Flows, $\mathbf{u}$, inferred from the advection of the supergranulation, plotted against latitude, λ . **a**, Flow in the direction of solar rotation, u_x (red). The green line shows the rotation of the small magnetic features¹⁰, and the black line is for the photospheric rotation⁵. The dotted line shows the pattern rotation obtained by tracking supergranular features with a 24-h delay, in agreement with an earlier measurement⁵. **b**, Northward meridional flow, u_y (red). The meridional flow of the magnetic features¹¹ (green) is again similar. The dotted line shows the anomalous results obtained by tracking the supergranulation pattern with a 24-h delay.

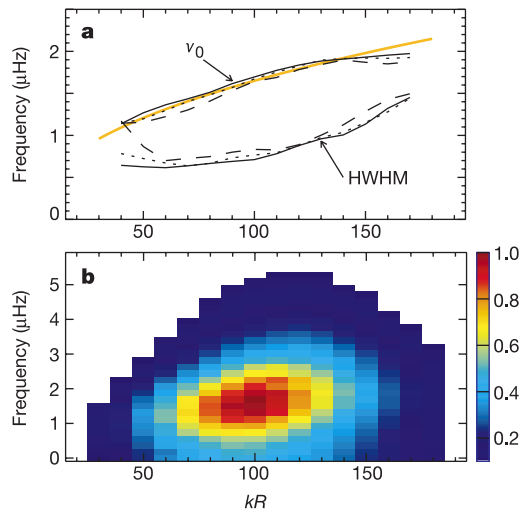


Figure 3 Average dynamical properties in a co-moving frame. **a**, Oscillation frequency ν_0 versus kR at latitudes $\lambda = 0^\circ$ (solid), $\lambda = \pm 25^\circ$ (dotted) and $\lambda = \pm 50^\circ$ (dashed). For reference, we plot the approximate dispersion relation $\nu = 1.65(kR/100)^{0.45} \mu\text{Hz}$ (orange). Also shown is the half-width at half-maximum (HWHM) of the lorentzian profiles for the same latitudes, implying an e-folding lifetime of 1–3 d. The quality factor, ν_0/HWHM , does not exceed 2. **b**, Power spectrum corrected for rotation and meridional circulation and averaged over azimuth and latitude. We note that the distribution of power as a function of frequency is affected only by the known temporal window function, whereas the wavenumber dependence includes effects of the telescope optics and of the time-distance analysis which are not fully understood.

relation $\nu = \nu_0(k)$. The waves have a rather low quality factor, as can be seen in the azimuthally averaged power spectrum (Fig. 3b).

As ν_0 and the dominant size¹² of supergranules are observed to be essentially independent of latitude, the general dynamics determining the timescale and the spatial scale of supergranulation are not affected by the Coriolis force associated with the large-scale vorticity (rotation). We observe, however, a pronounced anisotropy in the azimuthal distribution of wave power at fixed k (Fig. 4a). The power is maximum in the direction of rotation and towards the equator in both hemispheres (Fig. 4b). The pattern therefore senses the effect of rotation. A snapshot of the divergence field would not reveal this, as the sum of the powers measured in opposite directions is isotropic (Fig. 4a); the vorticity field, on the other hand, is known⁸ to be slightly sensitive to the effect of the Coriolis force.

Earlier estimates^{4,5} of the supergranulation rotation, obtained by tracking the supergranulation pattern from one image to the next, were systematically found to be higher¹³ than the rotation of the magnetic network (Fig. 2a). This apparent super-rotation of the pattern can now be understood, as waves are predominantly prograde. The east–west motion of the pattern is effectively a power-weighted average of the true rotation and the non-advective phase speed $u_w = 2\pi\nu_0/k \approx 65 \text{ m s}^{-1}$. Similarly, the excess of wave power towards the equator is reflected in the meridional motion of the pattern (Fig. 2b).

We have shown that supergranulation displays a high level of organization in space and time. It is known¹⁴ that heat and momentum transport in turbulent rotating convection is controlled by a network of coherent cyclonic plumes sinking from the thermal boundary layer. These plumes may be at the origin of the organization of the supergranulation pattern. We find that supergranulation supports waves. The prograde excess of wave power is probably due to the influence of rotation that breaks the east–west symmetry, allowing new instabilities to propagate. Recent numerical simulations¹⁵ of solar convection show convective patterns that move prograde relative to the local rotation at low latitudes, and may help explain the observations. Convection in oblique magnetic fields¹⁶

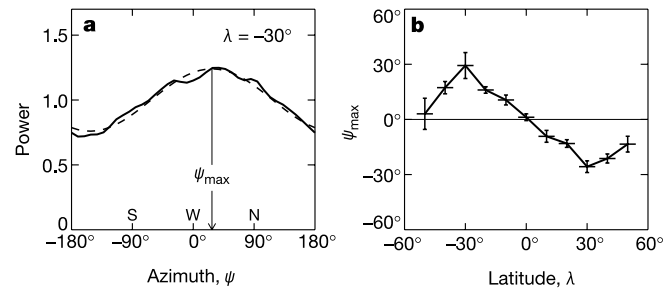


Figure 4 Wave power as a function of azimuth and latitude. **a**, Directional distribution of power at latitude $\lambda = -30^\circ$ (solid line), obtained by integrating $P_k(\nu, \psi)$ over frequencies $\nu > k\mathbf{u}/2\pi$ and then averaging over k . We applied an astigmatism correction estimated from the data. The azimuth of maximum power, ψ_{max} , is measured by fitting a sinusoidal function (dashed line). **b**, Plot of ψ_{max} versus latitude.

also exhibits solutions that take the form of travelling waves, where the tilt of the convection cells, their wave speed, and direction depend on the strength and obliquity of the field. Supergranulation appears to be a rare example of travelling wave convection in a very highly turbulent fluid, a nonlinear phenomenon that has been observed in laboratory experiments and numerical simulations^{15–18} under conditions of much weaker turbulence. We suggest that supergranular waves may be used as a diagnostic tool for probing the upper solar convection zone. □

Received 2 May; accepted 29 October 2002; doi:10.1038/nature01287.

- Leighton, R. B., Noyes, R. W. & Simon, G. W. Velocity fields in the solar atmosphere. I. Preliminary report. *Astrophys. J.* **135**, 474–499 (1962).
- Simon, G. W. & Leighton, R. B. Velocity fields in the solar atmosphere. III. Large-scale motions, the chromospheric network, and magnetic fields. *Astrophys. J.* **140**, 1120–1147 (1964).
- Galloway, D. J. & Weiss, N. O. Convection and magnetic fields in stars. *Astrophys. J.* **243**, 945–953 (1981).
- Duvall, T. L. Jr The equatorial rotation rate of the supergranulation cells. *Sol. Phys.* **66**, 213–221 (1980).
- Snodgrass, H. B. & Ulrich, R. K. Rotation of Doppler features in the solar photosphere. *Astrophys. J.* **351**, 309–316 (1990).
- Scherrer, P. H. *et al.* The solar oscillations investigation—Michelson Doppler imager. *Sol. Phys.* **162**, 129–188 (1995).
- Duvall, T. L. Jr *et al.* Time-distance helioseismology. *Nature* **362**, 430–432 (1993).
- Duvall, T. L. Jr & Gizon, L. Time-distance helioseismology with f modes as a method for measurement of near-surface flows. *Sol. Phys.* **192**, 177–191 (2000).
- Schou, J. & Bogart, R. S. Flow and horizontal displacements from ring diagrams. *Astrophys. J.* **504**, L131–L134 (1998).
- Komm, R. W., Howard, R. F. & Harvey, J. W. Rotation rates of small magnetic features from two- and one-dimensional cross-correlation analyses. *Sol. Phys.* **145**, 1–10 (1993).
- Komm, R. W., Howard, R. F. & Harvey, J. W. Meridional flow of small photospheric magnetic features. *Sol. Phys.* **147**, 207–223 (1993).
- Beck, J. G. *Large Scale Solar Velocities on Time Scales up to Thirty Days*. Thesis, Univ. California (1997).
- Beck, J. G. & Schou, J. Supergranulation rotation. *Sol. Phys.* **193**, 333–343 (2000).
- Brummell, N. H., Hurlburt, N. E. & Toomre, J. Turbulent compressible convection with rotation. I. Flow structure and evolution. *Astrophys. J.* **473**, 494–513 (1996).
- Miesch, M. S. *et al.* Three-dimensional spherical simulations of solar convection. I. Differential rotation and pattern evolution achieved with laminar and turbulent states. *Astrophys. J.* **532**, 593–615 (2000).
- Hurlburt, N. E., Matthews, P. C. & Proctor, M. R. E. Nonlinear compressible convection in oblique magnetic fields. *Astrophys. J.* **457**, 933–938 (1996).
- Zhong, F., Ecke, R. & Steinberg, V. Asymmetric modes and the transition to vortex structures in rotating Rayleigh–Bénard convection. *Phys. Rev. Lett.* **67**, 2473–2476 (1991).
- Walden, R. W. *et al.* Traveling waves and chaos in convection in binary fluid mixtures. *Phys. Rev. Lett.* **55**, 496–499 (1985).

Acknowledgements We thank D. O. Gough for suggestions about the general presentation of this Letter, and P. Milford, P. H. Scherrer, C. J. Schrijver and N. O. Weiss for comments. SOHO is a mission of international cooperation between the European Space Agency and NASA. MDI is supported by the Office of Space Sciences of NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to L.G. (e-mail: lgizon@solar.stanford.edu).

Volcanically emitted sodium chloride as a source for Io's neutral clouds and plasma torus

E. Lellouch*, G. Paubert†, J. I. Moses‡, N. M. Schneider§ & D. F. Strobel||

* Observatoire de Paris, F-92195 Meudon, France

† Institut de Radio-Astronomie Millimétrique, E-18012 Granada, Spain

‡ Lunar and Planetary Institute, Houston, Texas 77058-1113, USA

§ LASP, University of Colorado, Boulder, Colorado 80309-0392, USA

|| The Johns Hopkins University, Baltimore, Maryland 21218, USA

The atmosphere of Jupiter's satellite Io is extremely tenuous, time variable and spatially heterogeneous. Only a few molecules—SO₂, SO and S₂—have previously been identified as constituents of this atmosphere, and possible sources^{1–4} include frost sublimation, surface sputtering and active volcanism. Io has been known^{5,6} for almost 30 years to be surrounded by a cloud of Na, which requires an as yet unidentified atmospheric source of sodium. Sodium chloride has been recently proposed as an important atmospheric constituent, based on the detection of chlorine in Io's plasma torus^{7,8} and models of Io's volcanic gases⁹. Here we report the detection of NaCl in Io's atmosphere; it constitutes only ~0.3% when averaged over the entire disk, but is probably restricted to smaller regions than SO₂ because of its rapid photolysis and surface condensation¹⁰. Although the inferred abundance of NaCl in volcanic gases is lower than predicted⁹, those volcanic emissions provide an important source of Na and Cl in Io's neutral clouds and plasma torus.

Observations were performed on the nights of 14–18 January 2002 with the 30-m IRAM (Institut de Radio-Astronomie Millimétrique) radio telescope at Pico Veleta, Spain. The 42.5-h rotation period of Io allowed us to observe both sides of Io, covering each night ~50–80° of orbital longitude. We observed NaCl gas in two rotational transitions at wavelengths of 1.3 and 2.1 mm, simultaneously monitoring the temporally variable millimetre emission from SO₂. Both NaCl lines were detected each time they were searched for (Fig. 1), and at least three SO₂ lines were detected simultaneously (Table 1). The NaCl lines are detectable in integrations ~1.5-h long, but the signal-to-noise is not sufficient to search for orbital variability; here we focus on the analysis of the 15 and 17 January average (Fig. 2). Although the spatial resolution of the observations is only ~10 times the diameter of Io, the frequency match proves that these lines are associated with Io and not formed, for example, in the co-rotating plasma torus.

The NaCl lines, which represent thermal emission, exhibit characteristics similar to the SO₂ and SO millimetre lines previously detected on Io^{2,11,12}, with a linewidth of several hundreds of kilohertz and a contrast of a few tens of kelvin at Io. The simple consideration of the NaCl and SO₂ line contrasts at Io, in comparison to their intrinsic strengths¹³, demonstrates that NaCl is a minor compound with a typical mixing ratio relative to SO₂ of ~0.1%. But this estimate is crude, because it depends on the gas temperature and on the assumption that the NaCl and SO₂ lines have similar optical depths. In fact, the NaCl lines appear somewhat broader than their SO₂ counterparts (0.8–0.9 km s⁻¹ versus 0.5–0.6 km s⁻¹), indicating enhanced local NaCl mixing ratios and/or a higher NaCl temperature.

The detailed interpretation of the millimetre-wave lines from Io remains ambiguous, as their linewidth could reflect either thermal broadening in a hydrostatic atmosphere or velocity dispersion in a volcanic atmosphere². A hydrostatic and isothermal SO₂ atmo-

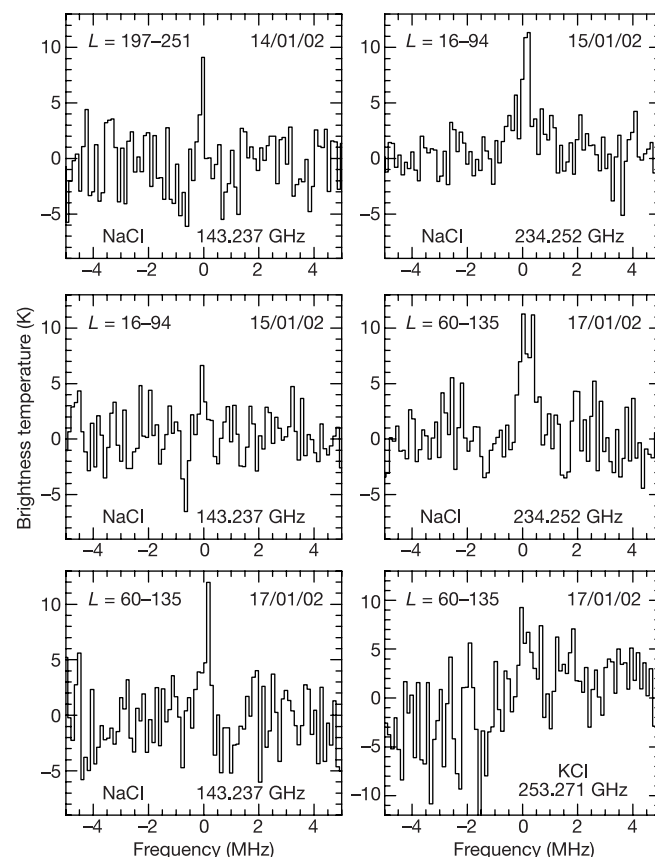


Figure 1 Overview of the NaCl observations. The NaCl line observations at 143.237 and 234.252 GHz are presented after smoothing to a resolution of 120 kHz. The frequency scale is the offset from rest frequency, after correction of Io's topocentric motion. Central meridian longitude (*L*) ranges are indicated. The bottom right panel shows tentative evidence for KCl at 253.271 GHz on 17 January, which, however, was not confirmed on the next night. The non-detection of KCl indicates an upper limit of ~1 for the KCl/NaCl ratio, consistent with the Na/K ratio in all types of chondrites (~13–17) and in Io's corona (10 ± 3 ; ref. 28).

sphere is described by its temperature T_{atm} , surface number density N_{SO_2} , and the fraction of the projected disk covered by the atmosphere (ρ_{SO_2}); these parameters were constrained from the SO₂ observations (see Methods section and Fig. 2 for details). The NaCl data were then analysed with a suite of models. Assuming NaCl to be co-located and at the same temperature as SO₂, the data imply a volume mixing ratio $\text{NaCl}/\text{SO}_2 = (2.5\text{--}5) \times 10^{-4}$. However, we favoured the case for denser NaCl atmospheres ($\text{NaCl}/\text{SO}_2 \approx 0.2\text{--}10\%$, 2σ error bars) covering a more restricted portion of the surface ($\rho_{\text{NaCl}} = 6\text{--}15\%$, versus $\rho_{\text{SO}_2} = 20\text{--}50\%$). The range of disk-averaged NaCl column density in this model is $8 \times 10^{12} - 2 \times 10^{14} \text{ cm}^{-2}$, with a preferred value of $4 \times 10^{13} \text{ cm}^{-2}$, that is, 0.35% of SO₂. Even better fits were found for higher NaCl gas temperatures (400–800 K) than for SO₂ (140–250 K).

In contrast to SO₂, sublimation equilibrium cannot be an immediate source of Io's NaCl atmosphere, given the extremely low vapour pressure of NaCl (~ 10^{-12} bar at 650 K, ref. 14). Sputtering of surface NaCl seems unlikely in the presence of a shielding SO₂ atmosphere of several times 10^{16} cm^{-2} (ref. 15). Although sputtering of salt-bearing atmospheric aerosols is not excluded, the most likely source seems to be direct volcanic output. Alkalis and chlorine are often present as condensates in terrestrial volcanic and fumarolic gases, and thermochemical equilibrium

Table 1 **IRAM 30-m observations**

Date*	UT range	Longitude	Species and frequencies (GHz)
2002 Jan. 14	22.10–04.30	197–251	SO ₂ (104.239, 143.057, 216.643, 265.482) NaCl (143.237)
2002 Jan. 15	19.15–04.30	16–94	SO ₂ (104.239, 143.057, 234.187, 265.482) NaCl (143.237, 234.252)
2002 Jan. 17	19.00–03.45	60–135	SO ₂ (104.239, 143.057, 234.187) NaCl (143.237, 234.252), KCl (253.271†)
2002 Jan. 18	18.20–23.55	259–306	SO ₂ (104.239), KCl (138.328†, 215.008†, 253.271†)

*UT date at beginning of observation.

† Undetected; all other lines in this table were detected.

calculations⁹ have predicted that NaCl is the dominant Na- and Cl-bearing molecular gas in Io volcanic emissions—this is the case over a wide range of quench temperature, vent pressure and gas bulk composition conditions. Photochemical models^{10,16} indicate that NaCl is subject to preferential photolytic and condensation losses in Io's atmosphere. Its lifetime against photolysis is about 3.5 h, compared to 50 h for SO₂. Condensation at the surface and on pre-existing aerosols further reduces its lifetime, although condensation within the atmosphere may not occur on a timescale shorter than the time of flight of a volcanic plume (typically 10–20 min). Thus, our detection of NaCl implies continuous active volcanism on Io at the time of the observations, and illustrates the importance of volcanism in feeding Io's atmosphere with gases other than O- and S-containing species. Unlike the case for SO₂, the NaCl lifetime is shorter than the typical hemispheric transport time (about 5 h, refs 17 and 18). This, and the likely role of sublimation sources for SO₂, would explain a NaCl atmosphere more localized than SO₂, perhaps restricted to ballistic distances from volcanic centres. The detection of gaseous NaCl also highlights its possible importance as a condensate on Io's surface^{9,10}.

To estimate the volcanic emission rates implied by our observations, we generated ballistic plume models, in which particles are ejected with fixed velocity of 0.5 km s⁻¹, and random direction^{2,12}. The SO₂ atmosphere is assumed to bounce twice after ballistic re-entry. The free parameters in this model are the flux of SO₂ molecules per volcano (Φ_{SO_2}) and the number of such volcanoes (N_{volc}) on one hemisphere. Assuming $T = 180$ K, a typical solution for the SO₂ lines is obtained with $\Phi_{\text{SO}_2} \approx 1.5 \times 10^{29}$ s⁻¹ and $N \approx 20$. This gives a total flux of 6×10^{30} SO₂ molecules s⁻¹, and a ~25% hemispheric coverage, consistent with inferences from Hubble Space Telescope Ly α observations¹⁹. If co-location with SO₂ is assumed, the NaCl lines can be explained with a total flux of $\Phi_{\text{NaCl}} = (2.2 \pm 0.5) \times 10^{27}$ NaCl molecules s⁻¹, that is, a $(4 \pm 1) \times 10^{-4}$ NaCl/SO₂ flux ratio. To explore the more plausible scheme of an increased NaCl concentration in the vicinity of vents, we alternatively assumed that the NaCl component of the plume does not bounce, mimicking complete condensation at the surface shock. In this case, the data require a higher NaCl flux of $(2-8) \times 10^{28}$ s⁻¹, that is, a 0.3–1.3% NaCl/SO₂ flux ratio.

Assuming bulk Na/S and Cl/S ratios of 0.05 and 0.04 in Io's erupting magmas, Fegley and Zolotov⁹ predicted a NaCl/SO₂ mixing ratio of 0.04 in high-temperature (>1,400 K) volcanic gas. This is several times (and perhaps up to an order of magnitude) higher than the above measured range. We note a 0.3–1.3% NaCl mixing ratio is consistent with a CI chondritic composition (having Na/S = 0.13 and Cl/S = 0.01), although higher Na/S and Cl/S ratios in Io's lithosphere may be expected owing to igneous differentiation²⁰. The problem becomes more severe if NaCl is co-located with SO₂ with a $\sim 4 \times 10^{-4}$ mixing ratio. This disagreement between observations and expectations may require a revision of our understanding of Io's lithospheric composition, although alternative explanations exist: (1) a significant fraction of the SO₂ atmosphere is sustained by sublimation equilibrium, (2) a loss process for NaCl

is missing in the thermochemical and/or photochemical models (for example, condensation at the vent), (3) NaCl is variable, whereas the observations represent a single 'snapshot'.

The NaCl detection provides clues to the origin of the sodium clouds that have been observed around Io for almost three decades^{1,5,6,21}. These include low-speed Na atoms (corona and B cloud), and several high-velocity (Na⁺) features for which the role of NaX⁺ ions has been demonstrated^{22,23}. Various production mechanisms

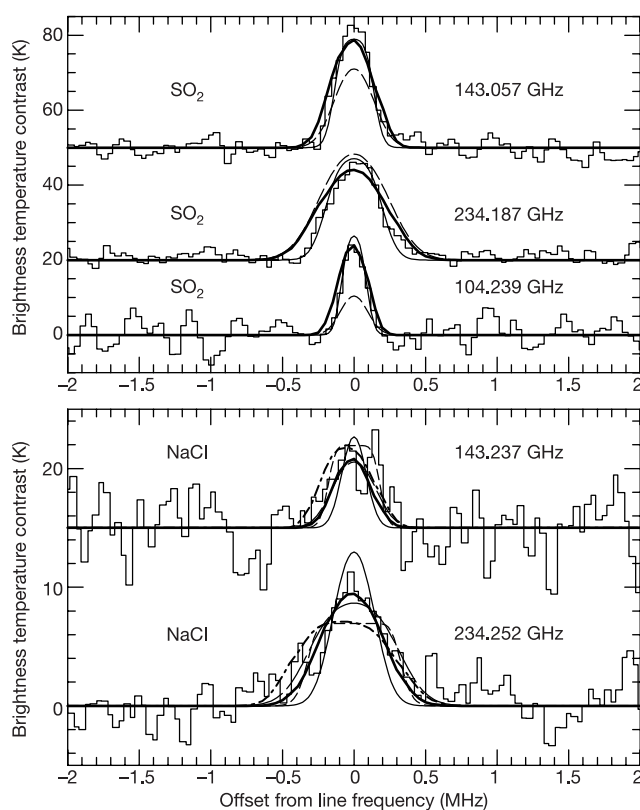


Figure 2 The SO₂ and NaCl 15 and 17 January observations, compared to various models. Histograms: observations, at 40 kHz resolution. Thin lines: 'hydrostatic' models. Thin solid lines: $T_{\text{atm}} = 180$ K, surface number density $N_{\text{SO}_2} = 2.5 \times 10^{10}$ cm⁻³, NaCl/SO₂ = 4×10^{-4} , projected fractional coverage $\rho_{\text{SO}_2} = \rho_{\text{NaCl}} = 32.2\%$. Thin short-dashed lines (in NaCl panel): NaCl/SO₂ = 1.5×10^{-2} , but $\rho_{\text{NaCl}} = 7.6\%$. Thin long-dashed line: $T_{\text{atm}} = 500$ K, $N_{\text{SO}_2} = 2 \times 10^{10}$ cm⁻³, NaCl/SO₂ = 5×10^{-3} , $\rho_{\text{SO}_2} = 12.3\%$, $\rho_{\text{NaCl}} = 2.5\%$. Thick lines: 'volcanic' models. Thick solid lines: model for 20 volcanoes on one hemisphere, each of them emitting $\Phi_{\text{SO}_2} = 1.5 \times 10^{29}$ SO₂ molecules s⁻¹ and $\Phi_{\text{NaCl}} = 5 \times 10^{25}$ NaCl molecules s⁻¹. Thick dashed-dotted lines (in NaCl panel): model for 20 volcanoes, each of them emitting $\Phi_{\text{NaCl}} = 1 \times 10^{27}$ NaCl molecules s⁻¹, but the NaCl atmosphere is assumed not to bounce (see main text and Methods section).

for NaX^+ have been discussed²⁴. For both slow and fast sodium, the supply rates are $\sim(0.5\text{--}2) \times 10^{26}$ atoms, and the inferred NaX^+ production rate is $(1\text{--}8) \times 10^{26} \text{ s}^{-1}$ (ref. 23).

A critical assessment of the recent discovery observations of Cl^+ and Cl^{2+} in the Io plasma torus^{7,8,25} indicates a Cl/S total ion ratio of $\sim 2\%$ (that is, a Cl/all-ion ratio of $\sim 1\%$). This is comparable to the torus Na/S ratio^{1,21,26}, suggesting that NaCl is the common parent of Na and Cl in Io's environment. However, photochemical modelling¹⁰ indicates that, even if molecular NaCl dominates in the lower atmosphere, atomic Na and Cl are respectively the major sodium- and chlorine-bearing species at the exobase; in this model, based on a hydrostatic equilibrium description, the NaCl upward flux at the top of the atmosphere is $\sim 10\text{--}20$ times smaller than the corresponding atomic Na and Cl fluxes, and $\sim 1,000$ times smaller than the upward NaCl flux at the bottom. Our observations indicate surface emission rates of $1.7 \times 10^{27}\text{--}8 \times 10^{28}$ NaCl molecules s^{-1} . Combining the two, the implied escape fluxes are $1.7 \times 10^{24}\text{--}8 \times 10^{25}$ NaCl molecules s^{-1} and $1.7 \times 10^{25}\text{--}1.6 \times 10^{27}$ Na and Cl atoms s^{-1} . While the NaCl escape is too small, the range for Na escape encompasses the NaX^+ production rate inferred from the fast sodium, as well as the supply rates for low-speed Na. We conclude that (1) the escape of atomic Na and Cl produced from the NaCl photolysis is an important source of chlorine and sodium in Io's environment, and (2) unless plume dynamics preferentially favour the escape of molecular NaCl, the production of fast sodium is dominated by reactions of torus molecular ions with atomic Na rather than by direct NaCl ionization. Improved measurements of neutral and ionized Na and Cl, the determination of the respective spatial distribution of NaCl and SO_2 , as well as a reassessment of escape rates in plume-like atmospheres, are now needed. □

Methods

Spectroscopy

We searched NaCl gas in two rotational transitions, at 143.237 GHz ($J = 11 \rightarrow 10$) and 234.252 GHz ($J = 18 \rightarrow 17$) respectively. Emission from SO_2 , routinely detectable but temporally variable², was monitored by observing simultaneously the SO_2 (16,2,14) $\rightarrow$ (16,1,15) and (28,3,25) $\rightarrow$ (28,2,26) lines at 143.057 and 234.187 GHz, respectively. Observing SO_2 and NaCl lines in the same receivers facilitated the interpretation in terms of the NaCl/ SO_2 relative abundance by minimizing calibration uncertainties. Additional SO_2 lines at 104.239, 216.482 and 265.482 GHz were observed (Table 1). Finally, KCl was unsuccessfully searched for at 253.272 and 215.008 GHz. The telescope beam size is about $10''$ at 234 GHz and $17''$ at 143 GHz.

Observations

All observations were conducted using a technique validated on previous millimetre-wave observations of Io (for example, refs 11, 12). Frequency tuning was checked on the interstellar source Ori A. The telescope pointing was determined by azimuth/elevation drifts across Jupiter, and checked every $\sim 1\text{--}2$ h. For velocity tracking, the telescope corrects for the topocentric motion of Jupiter but not for the jovicentric motion of Io. Therefore, the Io observations consisted of quick (2-min) scans, which, in the data reduction phase, were individually corrected for the jovicentric Doppler shift before co-addition, using ephemerides from IMCCE at (http://www.imcce.fr/ephemeride_eng.html). Observations were made in the 'wobbling switch' mode, in which, through chopping of the secondary at a frequency of 0.25 Hz, the telescope alternately points to the source and to a reference sky position 1–4 arcmin away. We used autocorrelator and filter-bank backends, providing spectral resolutions of 40–120 kHz. Sky conditions were excellent, with water amounts as low as 0.5–2 mm, resulting in zenith opacities of ~ 0.05 at 143 GHz and ~ 0.1 at 234 GHz.

Modelling

The SO_2 line observations were used to characterize the SO_2 atmosphere at the time of our observations. In a hydrostatic equilibrium, isothermal atmosphere model, the latter is described by its temperature T_{atm} , surface number density N_{SO_2} , and the fraction of the projected disk covered by the atmosphere (ρ_{SO_2}). All models assume an airmass of 1.5 and a brightness surface temperature of 88 K (ref. 27). The 14 and 15 January observations, which include the high-energy (580 K) 265.481 GHz SO_2 line, provide the best estimate of atmospheric temperature ($T_{\text{atm}} = 180 \pm 60$ K). Analysing the 15 and 17 January SO_2 data (3 lines) and exploring the (T_{atm} , N_{SO_2} , ρ_{SO_2}) parameter space with a χ^2 analysis, we obtained $T_{\text{atm}} = 140\text{--}250$ K, $N_{\text{SO}_2} = (1\text{--}5) \times 10^{10} \text{ cm}^{-3}$, $\rho_{\text{SO}_2} = 0.2\text{--}0.5$, and a best fit disk-averaged SO_2 column density of $1.1 \times 10^{16} \text{ cm}^{-2}$. The NaCl data were then analysed in a suite of models allowing fits of increased quality. If NaCl is assumed to be co-located and at the same temperature as SO_2 (that is, the same T_{atm} and ρ), the observations indicate a volume mixing ratio $\text{NaCl}/\text{SO}_2 = (2.5\text{--}5) \times 10^{-4}$, but the modelled NaCl lines

are too narrow (Fig. 2). We then relaxed the assumption of co-location of NaCl with SO_2 , while retaining the hypothesis of equal temperature. Significantly better fits were obtained (Fig. 2) for denser NaCl atmospheres ($\text{NaCl}/\text{SO}_2 \approx 0.2\text{--}10\%$, 2σ error bars) covering a more restricted portion of the surface ($\rho_{\text{NaCl}} = 6\text{--}15\%$ projected), although the 234.252 GHz line is now somewhat too broad. Because the two NaCl lines have similar optical depths, further discrimination between NaCl partial pressure and surface coverage is not possible. Finally, relaxing also the temperature constraint, an even better fit to the data is achieved if higher temperatures (400–800 K) and even lower surface coverages (1.5–7%) are used for NaCl gas than found for SO_2 (Fig. 2).

Simple volcanic models were explored by numerically generating ballistic plumes with a fixed (0.5 km s^{-1}) ejection velocity and random direction within a cone with 55° semi-angle². The SO_2 atmosphere was assumed to bounce twice after ballistic reentry, keeping each time 70% of its pre-shock energy. The gas temperature was assumed to be 180 K. The free parameters are the flux of SO_2 and NaCl molecules per volcano (Φ_{SO_2} and Φ_{NaCl}) and the number of such volcanoes, assumed identical, on one hemisphere. We explored models in which NaCl is co-located with SO_2 , and models in which the NaCl component of the plume does not bounce, simulating the complete NaCl surface condensation. The latter models are regarded to be more physically realistic.

Received 31 July; accepted 1 November 2002; doi:10.1038/nature01292.

- Spencer, J. R. & Schneider, N. M. Io on the eve of the Galileo mission. *Annu. Rev. Earth Planet. Sci.* **24**, 125–190 (1996).
- Lellouch, E. Io's atmosphere: not yet understood. *Icarus* **124**, 1–21 (1996).
- McGrath, M. A., Belton, M. J. S., Spencer, J. R. & Sartoretto, P. Spatially resolve spectroscopy of Io's Pele plume and SO_2 atmosphere. *Icarus* **146**, 476–493 (2000).
- Spencer, J. R. *et al.* Discovery of gaseous S_2 in Io's Pele plume. *Science* **288**, 1208–1210 (2000).
- Brown, R. A. in *Exploration of the Solar System* (eds Woszczyk, A. & Iwaniszewska, C.) 531–537 (Reider, Hingham, MA, 1974).
- Trafton, L., Parkinson, T. & Macy, W. The spatial extent of sodium emission around Io. *Astrophys. J.* **190**, L85–L89 (1974).
- Kueppers, M. & Schneider, N. M. Discovery of chlorine in the Io torus. *Geophys. Res. Lett.* **27**, 513–516 (2000).
- Feldman, P. D. *et al.* Detection of chlorine ions in the Far Ultraviolet Spectrum Explorer spectrum of the Io plasma torus. *Astrophys. J.* **554**, L123–L126 (2001).
- Fegley, B. Jr & Zolotov, M. Yu. Jr Chemistry of sodium, potassium, and chlorine in volcanic gases on Io. *Icarus* **148**, 139–210 (2000).
- Moses, J. L., Zolotov, M. Yu. & Fegley, B. Jr Alkali and chlorine photochemistry in a volcanically driven atmosphere on Io. *Icarus* **156**, 107–135 (2000).
- Lellouch, E. *et al.* The structure, stability, and global distribution of Io's atmosphere. *Icarus* **98**, 271–295 (1992).
- Lellouch, E. *et al.* Detection of sulfur monoxide in Io's atmosphere. *Astrophys. J.* **459**, L107–L110 (1996).
- Pickett, H. M. *et al.* Submillimeter, millimeter, and microwave spectral line catalog. *J. Quant. Spectrosc. Rad. Transfer* **60**, 883–890 (1998).
- Ewing, C. T. & Stern, K. H. Equilibrium vaporization rates and vapor pressures of solid and liquid sodium chloride, potassium chloride, potassium bromide, cesium iodide, and lithium fluoride. *J. Phys. Chem.* **78**, 1998–2005 (1974).
- Lanzetta, L. J. *et al.* Laboratory studies of charged particle erosion of SO_2 ice and applications to the frosts of Io. *Astrophys. J.* **259**, 920–929 (1982).
- Moses, J. L., Zolotov, M. Yu. & Fegley, B. Jr Photochemistry of a volcanically driven atmosphere on Io: sulfur and oxygen species from a Pele-type eruption. *Icarus* **156**, 76–106 (2000).
- Ingersoll, A. P., Summers, M. E. & Schlipf, S. G. Supersonic meteorology of Io: sublimation driven flow of SO_2 . *Icarus* **64**, 375–390 (1985).
- Summers, M. E. & Strobel, D. F. Photochemistry and vertical transport in Io's atmosphere and ionosphere. *Icarus* **120**, 290–316 (1996).
- Strobel, D. F. & Wolven, B. C. The atmosphere of Io: abundances and sources of sulfur dioxide and atomic hydrogen. *Astrophys. Space Sci.* **277**, 271–287 (2000).
- Kuskov, O. & Kronrod, V. A. Core sizes and internal structure of Earth's and Jupiter's satellites. *Icarus* **151**, 204–227 (2001).
- Thomas, N. Optical observations of Io's neutral clouds and plasma torus. *Surv. Geophys.* **13**, 91–164 (1992).
- Schneider, N. M. *et al.* Molecular origin for Io's fast sodium. *Science* **253**, 1394–1397 (1991).
- Wilson, J. K. & Schneider, N. M. Io's fast sodium: implications for molecular and atomic atmospheric escape. *Icarus* **111**, 31–44 (1994).
- Johnson, R. E. Formation of Na-containing molecular ions at Io. *Icarus* **111**, 65–72 (1994).
- Schneider, N. M., Park, A. H. & Kueppers, M. Spectroscopic studies of the Io torus during Galileo encounters: remote plasma diagnostics and the detection of Cl^{++} . *Bull. Am. Astron. Soc.* **32**, 35.03 (2000).
- Bagenal, F. & Sullivan, J. D. Direct plasma measurements in the Io torus and inner magnetosphere of Jupiter. *J. Geophys. Res.* **86**, 8447–8466 (1981).
- Altenhoff, W. J. *et al.* First radio astronomical estimate of the temperature of Pluto. *Astron. Astrophys.* **190**, L15–L17 (1988).
- Brown, M. E. Potassium in Europa's atmosphere. *Icarus* **151**, 190–195 (2001).

Acknowledgements We thank W. Thuillot for providing us with the topocentric velocities of Io necessary for the data reduction.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.L. (e-mail: emmanuel.lellouch@obspm.fr).

Implementation of the Deutsch–Jozsa algorithm on an ion-trap quantum computer

Stephan Gulde*, Mark Riebe*, Gavin P. T. Lancaster*, Christoph Becher*, Jürgen Eschner*, Hartmut Häffner*, Ferdinand Schmidt-Kaler*, Isaac L. Chuang*† & Rainer Blatt*

* Institut für Experimentalphysik, Universität Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria

† MIT Media Laboratory, Cambridge, Massachusetts 02139, USA

Determining classically whether a coin is fair (head on one side, tail on the other) or fake (heads or tails on both sides) requires an examination of each side. However, the analogous quantum procedure (the Deutsch–Jozsa algorithm^{1,2}) requires just one examination step. The Deutsch–Jozsa algorithm has been realized experimentally using bulk nuclear magnetic resonance techniques^{3,4}, employing nuclear spins as quantum bits (qubits). In contrast, the ion trap processor utilises⁵ motional and electronic quantum states of individual atoms as qubits, and in principle is easier to scale to many qubits. Experimental advances in the latter area include the realization of a two-qubit quantum gate⁶, the entanglement of four ions⁷, quantum state engineering⁸ and entanglement-enhanced phase estimation⁹. Here we exploit techniques^{10,11} developed for nuclear magnetic resonance to implement the Deutsch–Jozsa algorithm on an ion-trap quantum processor, using as qubits the electronic and motional states of a single calcium ion. Our ion-based implementation of a full quantum algorithm serves to demonstrate experimental procedures with the quality and precision required for complex computations, confirming the potential of trapped ions for quantum computation.

Laser-cooled trapped ions are ideally suited to the investigation and implementation of quantum information processing¹² because they exhibit these properties: (1) localization of the single particle to less than a few tens of nanometres^{13–15}; (2) control of the motional state down to the zero point of the trapping potential¹⁶; (3) a high degree of isolation from the environment and thus a very long time available for manipulations of their quantum state¹⁷; and (4) the ability to detect the ion’s quantum state with high precision by the electron shelving technique¹⁸. The same properties make single trapped ions well suited for storing quantum information in long-lived internal states¹⁹.

In our experiment we implement the Deutsch–Jozsa algorithm on a quantum processor based on a single trapped ⁴⁰Ca⁺ ion which is driven by laser pulses. A compensation technique for frequency shifts allows us to achieve the required control over the optical phases of the pulses²⁰. Following a recent proposal¹⁰, we also successfully combine ion-trap techniques for quantum state

manipulation with the method of composite pulses¹¹ adopted from NMR technology. Thus we achieve complete control over the ion’s motional and electronic state. The implementation of a quantum algorithm on an ion-trap processor, which we demonstrate here, serves as a test of the suitability of these techniques, particularly in view of their scalability towards a larger number of qubits.

To illustrate the Deutsch–Jozsa algorithm, we represent the four possible coins by four functions f that map one input bit ($a = 0, 1$ standing for ‘which side of the coin’) onto one output bit ($f(a) = 0, 1$ standing for ‘head or tail’). These functions can be divided into two constant functions $f_1(a) = 0, f_2(a) = 1$, representing the fake coins, and two balanced functions $f_3(a) = a, f_4(a) = \text{NOT } a$, which stand for the fair coins (see Table 1). An unknown function is characterized as constant or balanced by evaluating $f(0) \oplus f(1)$ which yields 0 (or 1) for a constant (or balanced) function ($\oplus$ denotes addition modulo 2). This evaluation classically requires two function calls, whereas the Deutsch–Jozsa quantum algorithm allows us to obtain the desired information with a single evaluation of the unknown f . The circuit diagram shown in Fig. 1 describes the implementation of the Deutsch–Jozsa algorithm with basic quantum operations²¹. The two qubits required for the Deutsch–Jozsa algorithm are encoded in the electronic state and in the phonon (vibrational quantum) number of the axial vibration mode of the single trapped ion (see Fig. 2). Qubit operations are realized by applying laser pulses on the ‘carrier’ or the ‘blue sideband’ of the electronic quadrupole transition as described in the Methods.

In general, a quantum algorithm is implemented by a sequence of such pulses on the carrier and sideband, but two major sources of error have to be overcome. First, as the simplest algorithms already require several pulses, we need to control precisely the relative optical phases of these pulses or, at least, to keep track of them such that the required pulse sequences lead to the desired operations. In particular, this requires the precise investigation and subsequent compensation of all phases introduced by the light shifts of the exciting laser beams. These light shifts arise as we have to drive

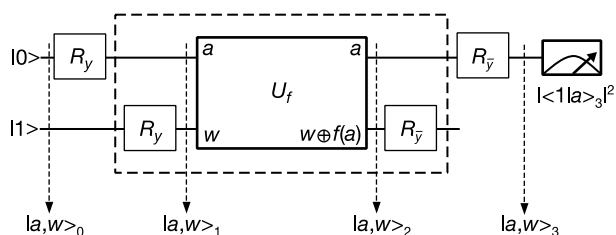


Figure 1 Quantum circuit for implementing the Deutsch–Jozsa algorithm with basic quantum operations. The upper line shows the input qubit $|a\rangle$ (‘which side of the coin’ information), the lower line an auxiliary working qubit $|w\rangle$ (corresponding to the channel on which the answer is provided). The rotations R_y (see Methods for details) create superpositions $|a_1\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $|w_1\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$ from the inputs $|a_0\rangle = |0\rangle$ and $|w_0\rangle = |1\rangle$. The box U_f represents a unitary operation specific to each of the functions f_n which applies f_n to a and adds the result to w modulo 2. Table 1 lists the logic operations required for transforming $|w\rangle$ into $|w \oplus f_n(a)\rangle$. The output of the box is $|a, w\rangle_2 = (|0, w_{in} \oplus f_n(0)\rangle + |1, w_{in} \oplus f_n(1)\rangle)/\sqrt{2}$. Up to an overall sign $|w\rangle$ is left unchanged, but the positive superposition $(|0\rangle + |1\rangle)/\sqrt{2}$ on $|a\rangle$ is transformed into a negative superposition $|a_2\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$ if f is balanced; otherwise it is unchanged. After the final rotations R_y , a measurement on $|a\rangle$ is performed with result $|a_3\rangle$ is either $|0\rangle$ or $|1\rangle$. Because of the sign change in $|a_2\rangle$ if f is balanced, $|\langle 1 | a_3 \rangle|^2 = f_n(0) \oplus f_n(1)$, that is, $|a_3\rangle$ yields the desired information whether the function f_n is balanced or constant. The working qubit w resumes its initial value $|w_3\rangle = |w_0\rangle = |1\rangle$.

Table 1 Truth table for the four possible functions

	Constant functions		Balanced functions	
	Case 1	Case 2	Case 3	Case 4
$f(0)$	0	1	0	1
$f(1)$	0	1	1	0
$w \oplus f(a)$	ID	NOT	CNOT	Z-CNOT

The third line is the effect of the logic function U_f on the qubit w : ID denotes the identity, CNOT is a controlled NOT operation, Z-CNOT is a zero controlled NOT, and the control bit in cases 3 and 4 is the input bit a .

sideband transitions (which couple much more weakly than carrier transitions) with high laser intensity. We cancel the unwanted light shifts with an additional off-resonant laser field, inducing a light shift of equal strength but opposite sign²⁰.

Second, a peculiarity of encoding a qubit within the ion's motional state is that we must ensure that the system does not leave the computational subspace $\{|S, 0_z\rangle, |D, 0_z\rangle, |S, 1_z\rangle, |D, 1_z\rangle\}$ (for notation, see Methods). The main problem here is that owing to the degenerate spectrum of a harmonic oscillator, sideband pulses work simultaneously on all levels. Therefore any population in $|S, 1_z\rangle$ prior to a blue sideband pulse will leave the computational subspace. To avoid this, we use composite pulses, that is, a sequence of carrier and/or sideband pulses that—up to an overall phase—constrain the system to the subspace¹⁰. We adopted this method from NMR technology¹¹. The translation of the Deutsch–Jozsa algorithm into composite pulses acting on the two qubits is described in the Methods.

For our experiments we load Ca ions into a linear Paul trap with axial frequency $\omega_z \approx 2\pi \times 1.7$ MHz. Figure 2 shows the relevant optical transitions used for laser cooling, state preparation and detection. Each experimental cycle starts with Doppler cooling for 2 ms on the $S_{1/2} \rightarrow P_{1/2}$ transition yielding average vibrational quantum numbers $\bar{n}_z \approx 20$. Further cooling of the axial motion to a ground state occupation of more than 99% is achieved by about 12 ms of sideband cooling⁸. To initialize the quantum processor in $|01\rangle = |S, 0_z\rangle$, we optically pump the ion to the $S_{1/2}$ ($m = -1/2$) state. Manipulations of both qubits are achieved by pulses from a stabilized titanium–sapphire laser (linewidth < 100 Hz, relative intensity noise $< 0.02_{\text{r.m.s.}}$) emitting at the $S_{1/2} \leftrightarrow D_{5/2}$ transition wavelength near 729 nm. In order to switch between R and R^+ rotations we shift the laser frequency with an acousto-optical modulator. The phase of the light field is switched via the phase of the radio frequency driving the acousto-optical modulator with an inaccuracy of less than 0.06 rad. Using the electron shelving technique⁸ we detect the ion's electronic state ($S_{1/2}$ or $D_{5/2}$) with a fidelity of 99.9% within a detection time of 3 ms.

We measure the fidelity of the implemented algorithm by repeating several thousand times the experimental sequence of cooling, initialization of both qubits, laser pulses for the algorithm and final measurement. Table 2 displays the achieved results. For cases 1, 3 and 4, the fidelity of identifying the function's class with a single measurement exceeds 97%; for case 2, it is above 90%. Note

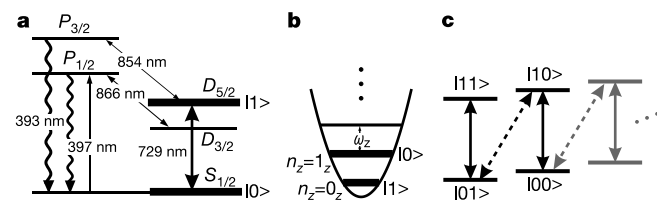


Figure 2 Quantum mechanical energy levels relevant for the ion-trap quantum computer. **a**, Ca^+ level scheme. The upper and lower electronic states $S_{1/2}$ ($m = -1/2$) and $D_{5/2}$ ($m = -1/2$) of the narrow quadrupole transition ($\tau_D \approx 1$ s) at 729 nm serve to implement one of the qubits, $|a\rangle$. Coherent radiation of a titanium–sapphire laser at 729 nm drives the qubit transition. Lasers at 397 nm, 866 nm and 854 nm are used for the excitation of resonance fluorescence, for Doppler cooling, and optical pumping. The laser system is described in detail elsewhere¹⁹. **b**, The lowest two number states, $n_z = 0, 1$, of the axial vibrational motion in the trap form the other qubit, $|w\rangle$. **c**, The combination of electronic states and energy eigenstates of the harmonic oscillator potential span the computational subspace. Numbers in ket notation denote the quantum logical values assigned to the respective states. Solid lines show carrier transitions; dashed lines show blue sideband transitions.

Table 2 Expected and measured results of the complete Deutsch–Jozsa algorithm

	Constant		Balanced	
	Case 1	Case 2	Case 3	Case 4
Expected $ \langle 1 a \rangle ^2$	0	0	1	1
Measured $ \langle 1 a \rangle ^2$	0.019(6)	0.087(6)	0.975(4)	0.975(2)
Expected $ \langle 1 w \rangle ^2$	1	1	1	1
Measured $ \langle 1 w \rangle ^2$	—	0.90(1)	0.931(9)	0.986(4)

The numbers in brackets are statistical 1σ uncertainties.

that to decide whether the function is constant or balanced, only $|\langle 1 | a \rangle|^2$ at the end of the algorithm needs to be measured. We also verified that the working qubit $|w\rangle$ is reset to its initial value by reading out the phonon number through a measurement of the Rabi frequency of the blue sideband transition^{8,16}.

The measured output of the algorithm shown in Table 2 slightly deviates from the ideal result. We identified the major sources for this infidelity and attribute it mainly to decoherence of the laser-atom phase, in particular caused by ambient magnetic field fluctuations²². Furthermore, in the implementation of case 2, which requires the most complex pulse sequence, we used higher laser power of the sideband transitions in order to speed up the algorithm and thus reduce the sensitivity to phase decoherence. This in turn caused off-resonant carrier excitation which limited the obtainable fidelity.

A major advantage of our state detection technique is the ability to follow the evolution of $|\langle 1 | a \rangle|^2$ during the quantum algorithm. For this, we truncate the pulse sequence at a certain time t and reveal $|\langle 1 | a(t) \rangle|^2$ by measuring the probability of finding the ion in the $D_{5/2}$ state. In Fig. 3 we display this probability as a function of time for all four cases. The data agree very well with the calculated ideal

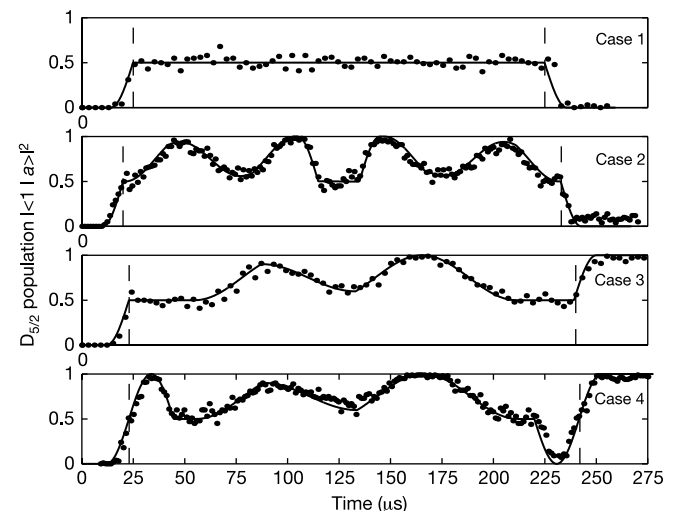


Figure 3 Time evolution of $|\langle 1 | a \rangle|^2$. Points are the probabilities, each inferred from 100 measurements, the line shows the ideal evolution. No parameters were adjusted to fit the data. The implementation of the functions $R_{y_w} U_{f_n} R_{y_w}$ takes place between the dashed lines. An initial R_{y_a} and a final R_{y_a} rotation on $|a\rangle$, implemented by carrier pulses, complete the algorithm. Taking case 3 as an example, R_{y_a} lasts from 12 μs to 22 μs . Then $R_{y_w} U_{f_n} R_{y_w}$ on $|a, w\rangle$ is implemented from 54 μs to 212 μs with the laser tuned to the blue sideband. The laser phase is switched at 87, 133 and 166 μs according to Table 3. The final R_{y_a} pulse is applied from 240 to 250 μs .

Table 3 **Implementations of $R_{y_w} U_f R_{y_w}$**

	Logic	Laser pulses
f_1	$R_{y_w} R_{y_w}$	No pulses
f_2	R_{y_w} SWAP ⁻¹ NOT _a SWAP R_{y_w}	$R^+(\frac{\pi}{2}, 0) R^+(\frac{2\pi}{\sqrt{2}}, \varphi_{\text{SWAP}}) R^+(\frac{\pi}{\sqrt{2}}, 0)$ $R(\frac{\pi}{2}, 0) R(\pi, \frac{\pi}{2}) R(\frac{\pi}{2}, \pi)$ $R^+(\frac{\pi}{\sqrt{2}}, \pi) R^+(\frac{2\pi}{\sqrt{2}}, \pi + \varphi_{\text{SWAP}}) R^+(\frac{\pi}{\sqrt{2}}, \pi)$ $R^+(\frac{\pi}{\sqrt{2}}, 0) R^+(\pi, \frac{\pi}{2}) R^+(\frac{\pi}{\sqrt{2}}, 0) R^+(\pi, \frac{\pi}{2})$
f_3	R_{y_w} CNOT R_{y_w}	
f_4	R_{y_w} Z-CNOT R_{y_w}	$R(\pi, 0) R^+(\frac{\pi}{\sqrt{2}}, 0) R^+(\pi, \frac{\pi}{2}) R^+(\frac{\pi}{\sqrt{2}}, 0) R^+(\pi, \frac{\pi}{2}) R(\pi, 0)$

The rotation angle for $R^+(\theta, \varphi)$ is given for the $|10\rangle \rightarrow |01\rangle$ transition. θ and φ denote the pulse duration and phase, respectively. $\varphi_{\text{SWAP}} = \arccos(\cot^2(\pi/\sqrt{2}))$, where the SWAP operation is explained in the Methods.

evolution (solid lines in Fig. 3, no fit parameters), demonstrating the high precision of the applied pulse sequence, especially the control over the optical phases.

The results demonstrate a high degree of control of all relevant experimental parameters, that is, laser frequency and intensity, optical phases, and trap frequency ω_z , over long pulse sequences. Therefore, the procedures presented here pave the way for implementing more complex algorithms and for scaling the system to multi-qubit operation. In particular, the light shift compensation technique demonstrated in this experiment can be directly transferred and advantageously applied to a several-qubit quantum processor. This technique will become increasingly important for scaling such a system because as the ion crystal becomes heavier, the higher laser intensities required to drive sideband transitions result in increased light shifts. Furthermore, by merging the composite pulse technique with our trapped-ion quantum computer we gain full access to all gate operations on the motional qubit. The employed composite-pulse phase gate also simplifies the Cirac–Zoller scheme⁵ for a universal set of quantum gates, by dispensing with the auxiliary level transition. Thus our procedures become applicable to a wider choice of ion species including $^{43}\text{Ca}^+$, which offers a potentially much longer coherence time than $^{40}\text{Ca}^+$. □

Methods

Encoding of qubits and single-qubit rotations

The two qubits required for the Deutsch–Jozsa algorithm are encoded in the electronic quantum state ($S_{1/2}$ ($m = -1/2$) $\equiv |0\rangle \equiv |S\rangle$ and $D_{5/2}$ ($m = -1/2$) $\equiv |1\rangle \equiv |D\rangle$) and in the phonon number of the axial vibration mode of the single trapped ion ($n_z = 0_z \equiv |1\rangle$ and $n_z = 1_z \equiv |0\rangle$). Note the counterintuitive encoding of the vibrational mode, which simplifies the desired initial state preparation in $|01\rangle = |S, 0_z\rangle$. The operations which modify the electronic qubit ('single-qubit rotations') are performed with laser pulses on the carrier ($|S, n_z\rangle \leftrightarrow |D, n_z\rangle$) transition, that is, no change of vibrational quantum number, laser on resonance. To connect the two qubits ('two-qubit rotations') the laser is detuned by $+\omega_z$ from the $|S\rangle \leftrightarrow |D\rangle$ resonance to the 'blue sideband' ($|S, n_z\rangle \leftrightarrow |D, n_z + 1\rangle$) as indicated in Fig. 2. Qubit rotations can be written as unitary operations in the following way¹²:

Carrier rotations are given by

$$R(\theta, \phi) = \exp \left[i \frac{\theta}{2} (e^{i\phi} \sigma^+ + e^{-i\phi} \sigma^-) \right]$$

whereas transitions on the blue sideband are denoted as

$$R^+(\theta, \phi) = \exp \left[i \frac{\theta}{2} (e^{i\phi} \sigma^+ b^\dagger + e^{-i\phi} \sigma^- b) \right]$$

Here $\sigma^\pm$ are the atomic raising and lowering operators which act on the electronic quantum state of the ion, that is, the first qubit, by inducing transitions from the $|S\rangle$ to $|D\rangle$ state and vice versa (notation: $\sigma^+ = |D\rangle\langle S|$). The operators b and $b^\dagger$ stand for the annihilation and creation of a phonon at the trap frequency, that is, they work on the motional quantum state, the second qubit. The parameter θ depends on the strength and the duration of the applied pulse and ϕ is its phase, that is, the relative phase between the optical field and the atomic polarization. We use the definitions $R_y = R(\pi/2, 0)$ and $R_z = R(\pi/2, \pi)$.

Translation of the Deutsch–Jozsa algorithm into composite pulses

The quantum circuit shown in Fig. 1 shows the quantum logic operations used for the

implementation and Table 1 lists the logic functions corresponding to the unitary operations U_{f_i} . The R_y rotations on the electronic qubit $|a\rangle$ are carrier pulses. For efficient computation we combine the rotations R_y, R_y on $|w\rangle$ and the manipulations for implementing U_f into an optimized pulse sequence, $R_{y_w} U_{f_i} R_{y_w}$ (dashed box in Fig. 1). As these operations act also on the motional state, we implement them with pulses on the carrier and the blue axial sideband. However, sideband pulses operate on both qubits simultaneously. Thus, for operations on $|w\rangle$ alone, we first swap the information from $|w\rangle$ into $|a\rangle$ with a sequence of three blue sideband pulses, then we rotate $|a\rangle$ as desired and swap back.

For a swap operation one might be tempted to use a single π -pulse on the blue sideband. However, applying this to the state $|00\rangle = |S, 1_z\rangle$ leads to a population of states with two phonons outside the computational subspace. Therefore we use a composite pulse sequence consisting of three pulses, whose lengths are chosen such that starting from $|S, 1_z\rangle$ the ion is rotated by $\pi, 2\pi$ and π , respectively. As a result the ion is rotated by 4π back to $|S, 1_z\rangle$ independently of the pulses' relative phases. In addition, using the blue sideband ensures that $|11\rangle = |D, 0_z\rangle$ also stays unchanged as required for the swap operation.

The desired swap operation $|S, 0_z\rangle \leftrightarrow |D, 1_z\rangle$ is possible because compared to the $|S, 1_z\rangle \leftrightarrow |D, 2_z\rangle$ transition, the Rabi frequency for the $|S, 0_z\rangle \leftrightarrow |D, 1_z\rangle$ transition is smaller by $1/\sqrt{2}$ (refs 8, 16). So in this manifold the three pulses' lengths correspond to rotation angles of $\pi/\sqrt{2}, 2\pi/\sqrt{2}, \pi/\sqrt{2}$. It can be shown that choosing the laser-atom phase of the second pulse to be $\arccos(\cot^2(\pi/\sqrt{2})) = \pi \cdot 0.3033 \dots$ relative to the first and the third pulses, the populations of $|10\rangle = |D, 1_z\rangle$ and $|01\rangle = |S, 0_z\rangle$ are exchanged. This realises the desired swap. Table 3 (case 2) lists the complete pulse sequence for the implementation of $R_{y_w} U_{f_i} R_{y_w}$. Similar procedures are applied to realise the pulse sequences for cases 3 and 4. In these cases the rotations R_{y_w}, R_{y_w} and the operations required for U_{f_i}, U_{f_i} can be combined in such a way that swap operations become unnecessary.

Received 21 October; accepted 27 November 2002; doi:10.1038/nature01336.

- Deutsch, D. Quantum theory, the Church–Turing principle and the universal quantum computer. *Proc. R. Soc. Lond. A* **400**, 97–117 (1985).
- Deutsch, D. & Jozsa, R. Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A* **439**, 553–558 (1992).
- Chuang, I. I., Vandersypen, I. M. K., Zhou, X., Leung, D. W. & Lloyd, S. Experimental realization of a quantum algorithm. *Nature* **393**, 143–146 (1998).
- Jones, T. F. & Mosca, M. Implementation of a quantum algorithm to solve Deutsch's problem on a nuclear magnetic resonance quantum computer. *J. Chem. Phys.* **109**, 1648–1653 (1998).
- Cirac, J. I. & Zoller, P. Quantum computations with cold trapped ions. *Phys. Rev. Lett.* **74**, 4091–4094 (1995).
- Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
- Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
- Roos, Ch. *et al.* Quantum state engineering on an optical transition and decoherence in a Paul trap. *Phys. Rev. Lett.* **83**, 4713–4716 (1999).
- Meyer, V. *et al.* Experimental demonstration of entanglement-enhanced rotation angle estimation using trapped ions. *Phys. Rev. Lett.* **86**, 5870–5873 (2001).
- Childs, A. M. & Chuang, I. M. Universal quantum computation with two-level trapped ions. *Phys. Rev. A* **63**, 012306 (2001).
- Levitt, M. H. Composite pulses (NMR spectroscopy). *Prog. Nucl. Magn. Reson. Spectrosc.* **18**, 61–122 (1986).
- Šašura, M. & Bužek, V. Cold trapped ions as quantum information processors. *J. Mod. Opt.* **49**, 1593–1647 (2002).
- Eschner, J., Raab, Ch., Schmidt-Kaler, F. & Blatt, R. Light interference from single atoms and their mirror images. *Nature* **413**, 495–498 (2001).
- Guthöhrlein, G. R., Keller, M., Hayasaka, K., Lange, W. & Walther, H. A single ion as a nanoscopic probe of an optical field. *Nature* **414**, 49–51 (2001).
- Mundt, A. B. *et al.* Coupling a single atomic quantum bit to a high finesse optical cavity. *Phys. Rev. Lett.* **89**, 103001 (2002).
- Meekhof, D. M., Monroe, C., King, B. E., Itano, W. M. & Wineland, D. J. Generation of nonclassical motional states of a trapped atom. *Phys. Rev. Lett.* **76**, 1796–1799 (1996).
- Monroe, C., Meekhof, D. M., King, B. E. & Wineland, D. J. A "Schrödinger Cat" superposition state of an atom. *Science* **272**, 1131–1136 (1996).
- Dehmelt, H. Proposed $10^{14} \Delta\nu > \nu$ laser fluorescence spectroscopy on Ti^+ mono-ion oscillator. *Bull. Am. Phys. Soc.* **20**, 60 (1975).
- Nägerl, H. C. *et al.* Investigating a qubit candidate: Spectroscopy on the $S_{1/2}$ to $D_{5/2}$ transition of a trapped calcium ion in a linear Paul trap. *Phys. Rev. A* **61**, 023405 (2000).
- Häffner, H. *et al.* Precision measurement and compensation of optical Stark shifts for an ion-trap quantum processor. Preprint available at (<http://arXiv.org/abs/physics/0212040>) (2002).
- Nielsen, M. A. & Chuang, I. J. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2000).
- Schmidt-Kaler, F. *et al.* Coherence of qubits based on single Ca ions. Preprint available at (<http://arXiv.org/abs/quant-ph/0211059>) (2002).

Acknowledgements We gratefully acknowledge support by the European Commission (QSTRUCT, QI, QUEST and QUBITS networks), by the Austrian Science Fund (FWF), and by the Institut für Quanteninformation GmbH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.S.-K. (e-mail: Ferdinand.Schmidt-Kaler@uibk.ac.at).

Quasi-phase-matched generation of coherent extreme-ultraviolet light

A. Paul*, R. A. Bartels*, R. Tobey*, H. Green*, S. Weiman*, I. P. Christov†, M. M. Murnane*, H. C. Kapteyn* & S. Backus*

* Department of Physics and JILA, University of Colorado and National Institute of Standards and Technology, Boulder, Colorado 80309-0440, USA

† Department of Physics, Sofia University, Sofia 1164, Bulgaria

High-harmonic generation is a well-known method of producing coherent extreme-ultraviolet (EUV) light, with photon energies up to about 0.5 keV (refs 1, 2). This is achieved by focusing a femtosecond laser into a gas, and high harmonics of the fundamental laser frequency are radiated in the forward direction^{3,4}. However, although this process can generate high-energy photons, efficient high-harmonic generation has been demonstrated only for photon energies of the order 50–100 eV (ref. 5). Ionization of the gas prevents the laser and the EUV light from propagating at the same speed, which severely limits the conversion efficiency. Here we report a technique to overcome this problem, and demonstrate quasi-phase-matched frequency conversion of laser light into EUV. Using a modulated hollow-core waveguide to periodically vary the intensity of the laser light driving the conversion, we efficiently generate EUV light even in the presence of substantial ionization. The use of a modulated fibre shifts the energy spectrum of the high-harmonic light to significantly higher photon energies than would otherwise be possible. We expect that this technique could form the basis of coherent EUV sources for advanced lithography and high-resolution imaging applications. In future work, it might also be possible to generate isolated attosecond pulses.

Advances in nonlinear optics have greatly expanded the utility of the laser⁶. Using nonlinear crystals and frequency-conversion techniques such as frequency doubling⁷ and quasi-phase matching (QPM)^{8–10}, laser light at one wavelength can be converted to another. In nonlinear optics, efficient conversion requires that the process be phase matched⁶. As a pump beam propagates through a medium, the nonlinear response leads to generation of a signal beam at a different wavelength. If the two waves travel with the same phase velocity, the nonlinear response continues to add constructively to the signal beam, leading to a bright, 'phase-matched' output signal at a new wavelength. If a process cannot be phase matched, the phase between the signal field and the nonlinear response will eventually reach 180°, resulting in destructive interference and back-conversion into the fundamental field. The propagation distance at which destructive interference occurs is defined as the coherence length.

Phase matching in the visible is often accomplished using anisotropic crystals. Here, the signal and pump beams propagate with different polarizations, taking advantage of the birefringence of the medium to equalize propagation velocities. However, other approaches are possible. In 1962, shortly after the birth of nonlinear optics, Armstrong *et al.*¹¹ proposed "phase corrective schemes", later termed QPM schemes. In QPM, a phase mismatch is periodically 'corrected', with a periodicity corresponding to twice the coherence length. For example, by reversing the crystal orientation of the medium, the polarity of the nonlinear response is reversed just before destructive interference occurs. However, practical implementation of this concept awaited the development of crystal-poling techniques in the mid-1990s^{8–10}.

Unfortunately, reliance on solid materials for nonlinear frequency conversion has severely limited nonlinear optical techniques for generating very-short-wavelength light. Most solid materials are transparent only at wavelengths greater than 150–200 nm. Thus, the

generation of coherent light at EUV wavelengths must take place in a gas^{3,4}, where traditional phase-matching techniques do not work. Nevertheless, phase-matched frequency conversion to the EUV is still possible, and leads to the generation of fully spatially¹² and temporally¹³ coherent beams. In past work, we demonstrated the generation of phase-matched high-harmonic generation (HHG) over an extended propagation range, by confining the light in a gas-filled hollow waveguide^{5,14}. The effect of the neutral gas index of refraction on the pump light (decreasing the phase velocity of the pump) can be counterbalanced by waveguide dispersion and ionization (increasing the phase velocity of the pump), making it possible to equalize the propagation velocity of the pump and signal. However, this approach to phase matching is limited to relatively low EUV photon energies. Higher EUV photon energies are generated at higher levels of ionization, where the phase velocity of the pump beam is too fast to be phase matched using any technique demonstrated to date. Thus, although HHG can generate photons with energy of about 500 eV (refs 1, 2), efficient phase-matched HHG has been demonstrated only at photon energies of ≤ 50 –90 eV. This prevents the use of this source for applications such as *in vitro* imaging of small cellular structures, which requires EUV light in the 'water window' (~ 300 eV) region of the spectrum. Coherent sources at the 13-nm wavelength of interest to EUV lithography would also benefit from increased efficiency.

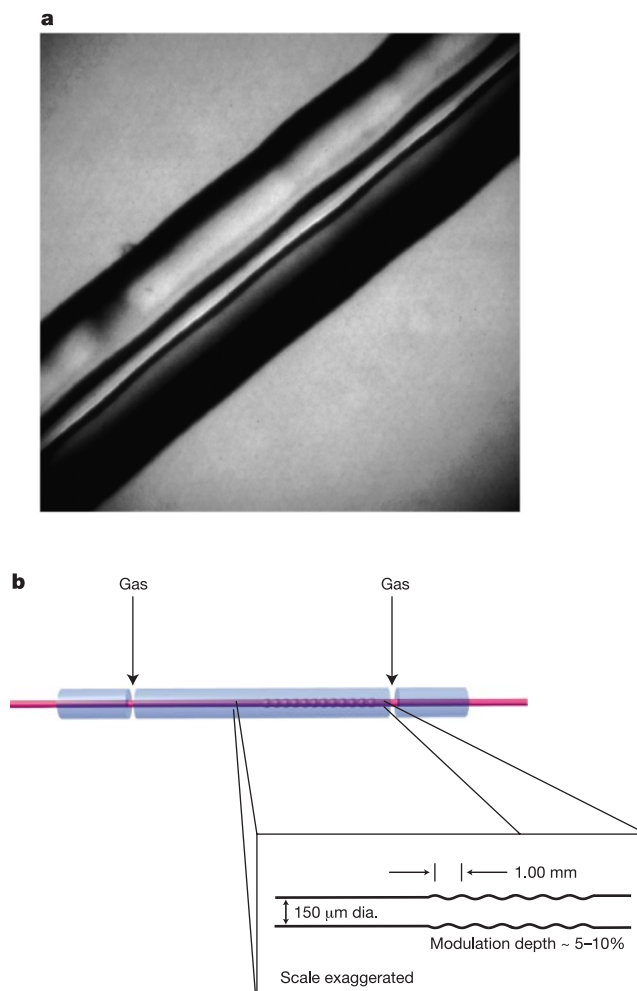


Figure 1 Modulated hollow-core fibres used to implement quasi-phase matching in the EUV. **a**, Picture of hollow-core modulated fibre, **b**, Details of modulated fibre with periodicity of 1 mm, inner diameter of 150 μm , and modulation depth of 10 μm .

Here we demonstrate that this limitation can be overcome by applying QPM techniques to HHG. Using a hollow-core waveguide with a periodically modulated diameter, the energy range over which we efficiently generate high harmonics can be increased significantly. The periodic structure modulates the generation of high harmonics and restricts HHG to regions with favourable constructive interference. It may also allow for the generation of isolated attosecond pulses. This work shows that sophisticated concepts of nonlinear optical 'photonics' and engineered structures can be applied even to the extreme nonlinear optics of HHG. It also complements our past work using optimally shaped light pulses to selectively optimize HHG¹³. Both works show that the highly nonlinear nature of the HHG process, rather than being a limitation, in fact allows for new control mechanisms that are simply not possible in conventional nonlinear optics.

In our experiment, 25-fs-duration pulses from a high-repetition-rate (2–5 kHz, 1 mJ per pulse) Ti:sapphire laser system operating at 760 nm (ref. 15) were focused into 150- μ m-diameter hollow fibres

filled with various gases. Some of the waveguides were periodically modulated in diameter to modulate the laser intensity inside the fibre. The modulated fibres used in this experiment were produced using precision glass-blowing techniques starting with a straight hollow-core fibre (also 150 μ m in diameter). The modulations were approximately sinusoidal, with a period of 1–0.5 mm, and a radial depth of the order of 10 μ m, corresponding to a 13% modulation of the fibre radius (Fig. 1).

Figure 2 shows the phase-matched HHG spectrum from He, Ne and Ar gas for straight (blue) and modulated (red) fibres. For all cases, the flux measured from the QPM fibres is significantly greater (by factors of 2–5) than that measured from straight fibres. Most significantly, the modulated fibre increases the brightness of higher harmonic orders by at least two orders of magnitude. In the case of He, for the straight fibre the HHG comb spans an energy range from 60 to 80 eV, with a spectral peak at 68 eV (Fig. 2a, blue trace). For the modulated fibre, the spectral peak shifts by 27 eV, from 68 to 95 eV (Fig. 2a, red), while the QPM harmonic comb also spans a broader energy range from 63 to 112 eV. Note that the wavelength region over which we achieve QPM easily includes the 94-eV (13 nm wavelength) region of interest for the next generation of EUV lithographies. In the case of Ne, use of a modulated fibre extends the high-flux harmonic emission from 70 to 90 eV, while significantly increasing the flux (Fig. 2b). In the case of Ar, the HHG spectrum from a straight fibre consists of a comb of phase-matched harmonics, peaked around 37 eV (Fig. 2c, blue). For the modulated fibre (Fig. 2c, red), the HHG output spectral peak shifts to 47 eV, while the flux increases significantly. We also investigated the HHG emission from Ar when the laser intensity was reduced by a factor of three. In this case (Fig. 3), the characteristic 'harmonics' merge into an almost smooth continuum of both even and odd orders. Moreover, the spectral bandwidth of the merged peaks for the QPM case corresponds to a single 250-as-duration EUV pulse, assuming a time-bandwidth limit (Fig. 3 inset).

We also investigated HHG from longer (2.5 cm) modulated fibres with different modulation periods. Figure 4 shows the experimen-

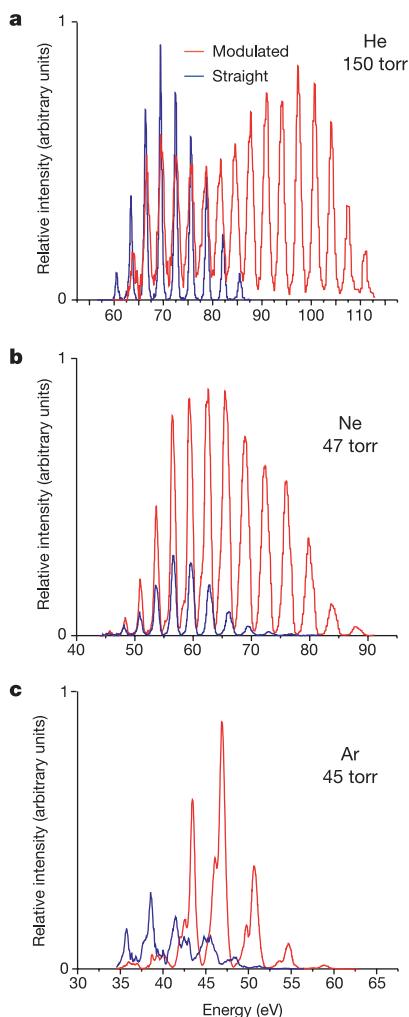


Figure 2 Experimentally measured HHG spectra for straight (blue) and modulated (red) fibres. Data are shown for He gas (a), Ne gas (b) and Ar gas (c), at pressures of 150 torr, 47 torr and 45 torr, respectively. In this case, the modulated section was 1 cm in length, with 1 mm periodicity, placed near the end of the fibre. The measured flux (non-optimized) corresponds approximately to 1 nJ per harmonic per pulse for Ar, and 20 pJ per harmonic per pulse for He at repetition rates of 2 kHz. These numbers are expected to increase when sufficient laser intensity is available to push the HHG spectra into the transparency region of the gas, and when the lengths of the modulated fibres are optimized.

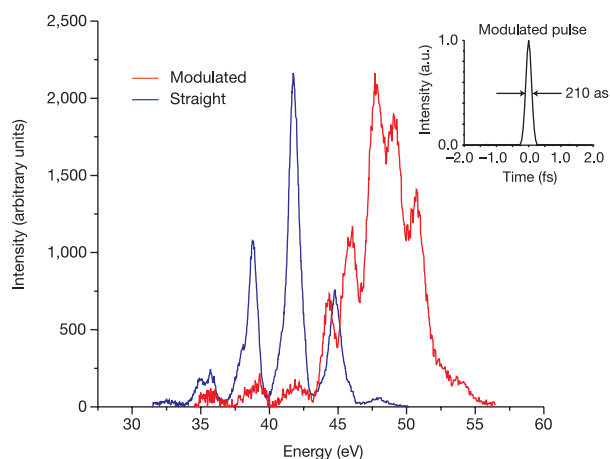


Figure 3 Experimentally measured HHG spectra from Ar for straight (blue) and modulated (red) fibres, at lower intensities than Fig. 2c, and at a pressure of 25 torr. In this case, the modulated section was 1 cm in length, with 1 mm periodicity, placed in the centre of the fibre. Inset, Fourier transform of the HHG emission from the modulated fibre, predicting the generation of a 250-as pulse, provided that the phase is flat. The reason for the different characteristic QPM spectra from Ar in Fig. 2c and Fig. 3 results from the use of different laser intensities and gas pressures in the two cases, resulting in a different temporal phase-matching window. Thus, the effect of QPM is evident in both cases, and has the potential for generating isolated attosecond pulses under the correct conditions. a.u., arbitrary units.

tally measured HHG spectra from He for three different periodicities of the modulated fibres. As the modulation period is reduced from 1 mm, to 0.75 mm, to 0.5 mm, the high-energy cut-off increases from 112 to 175 eV. This cut-off is limited by the available laser intensity, and is expected to increase significantly when longer modulated sections, lower pressures, shorter laser pulses, or higher laser intensities are used.

To understand our results, we use previous theoretical predictions of the potential utility of QPM frequency conversion in the EUV^{16,17}. HHG is extremely sensitive to intensity; thus a periodic modulation in the intensity of the driving laser will modulate the HHG emission. Generation of the highest harmonic orders is 'turned off' when the fibre bulge increases the mode diameter in the fibre, thus preventing back-conversion of the EUV light. Therefore HHG will be quasi-phase matched when the period of the modulation matches the period of the phase mismatch between the pump and signal (that is, twice the coherence length). We also note that two other experimental works used QPM in low-order harmonic generation in gases^{18,19}. However, their approach is not applicable to the EUV. Other proposed schemes, such as suppressing HHG with interfering beams, may be applicable in the EUV²⁰.

An intuitive picture of our experimental results can be obtained in the perturbative approach. The signal corresponding to the q th harmonic ($E_q(L)$) after propagation through a generation medium of length L is given by:

$$E_q(L) \approx \int_0^L E_\omega^q(z) d(z) e^{i\Delta k z} dz \quad (1)$$

where $E_\omega(z)$ is the laser field, $d(z)$ is the effective nonlinear coefficient for HHG, and $\Delta k = qk_\omega - k_{q\omega}$ is the net phase mismatch between the fundamental and harmonic field. In the absence of phase matching or quasi-phase matching, the rapid $e^{i\Delta k z}$ phase term will cause the integral in equation (1) to average to zero. This integral will be non-zero if either $\Delta k = 0$, corresponding to traditional phase matching in the visible or EUV, or if $d(z) \approx \cos(\Delta k_M z)$, corresponding to traditional QPM in the visible. For

the latter, the sign on the nonlinear coefficient $d(z)$ is modulated periodically using periodically poled materials^{8–10,21}. The modulation period Λ (where $\Delta k_M = 2\pi/\Lambda$) is chosen to be equal to the coherence length of the signal, that is, the distance over which the fundamental and harmonic fields undergo a phase slip of π , to ensure that the harmonic signal from different regions in the interaction length always adds in phase.

In our work, a new type of QPM can be implemented in the EUV where the signal wave itself is modulated periodically, that is, $E_\omega^q(z) \approx \cos(\Delta k_M z)$ and d is independent of z . The periodic QPM structure can modulate the generation of high harmonic orders in several ways. Harmonics near cut-off that require higher laser intensity for generation can be turned on and off by the modulation. However, other effects (such as the periodic evolution of the phase of the driving pulse) can also contribute to a periodic modulation of the harmonic generation process. In fact, virtually any periodic change in the character of the harmonic emission can allow QPM to operate, particularly in a very-high-order nonlinear process where very small phase changes can dramatically change the output.¹³

The significance of the quasi-phase matching discussed here is that it permits phase matching of HHG at higher ionization levels and hence higher photon energies than previously possible. This can be made apparent by calculating the coherence length that results from ionization of the gas. The plasma-induced change in index of refraction corresponds to a phase mismatch in equation (1) of $\Delta k_{\text{plasma}} \approx qn_e e^2 \lambda / 4\pi m_e \epsilon_0 c^2$, where λ is the laser wavelength, e is the charge of the electron, m_e is the mass of the electron, ϵ_0 is the vacuum permittivity, and n_e is the electron density. For fully ionized argon at a pressure of 1 torr, and for $q = 29$, $n_e = 3.3 \times 10^{16} \text{ cm}^{-3}$, giving $\Delta k \approx 23 \text{ cm}^{-1}$. Therefore, the coherence length, L_c , given by $L_c = \pi/\Delta k$, is 1.4 mm. Thus, very substantial levels of ionization can be compensated for using the QPM technique demonstrated here, with experimentally realizable modulation periods in the millimetre range. In our experiments, because we can modulate at most a 2.5-cm-long section of fibre, we used higher pressures (30 torr). This implies that the 1 mm modulation period compensates approximately for an additional 4% fractional ionization at harmonic order 29. This is to be compared with a straight fibre where phase matching occurs at ionization levels of 2–4%; thus, QPM more than doubles the allowable level of ionization in this case. For He, the effect of the QPM is greater, because the fractional ionization at which high harmonics are efficiently generated is much lower (<0.6% for $q = 61$, for example). Therefore, a small change in allowable ionization results in a large change in the energy of the highest phase-matched harmonic order. Finally, by decreasing the modulation period from 1 to 0.5 mm (Fig. 4), the amount of ionization that can be compensated increases, allowing for even higher harmonic orders to be phase matched. Furthermore, by increasing the length of the modulated fibre and reducing the pressure, very high levels of ionization might be compensated for, leading to the generation of very-high-order harmonics, perhaps even from fully ionized gasses. This approach also leads to a significant increase in output flux, because the transparency of the gas is increasing with increasing photon energy over the range we investigate, while the effective nonlinearity is not changing rapidly. For He and Ne for example, the absorption length is increasing with increasing harmonic photon energies. Therefore, we observe both higher harmonic orders and increased overall flux in the data of Fig. 2.

The QPM structure also suggests the generation of isolated attosecond pulses, even though the driving pulse is nearly two orders of magnitude longer than the generated pulse. This behaviour results from the fact that a dynamic, plasma-induced phase dominates the phase-matching conditions. As a laser beam propagates through a region of space, the level of ionization increases monotonically with time. At a critical ionization level, the phase-matching conditions will be optimal, and the bulk of the observed

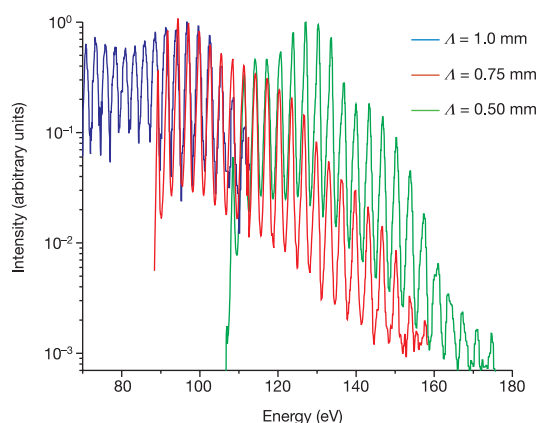


Figure 4 Experimentally measured HHG spectra (log scale) from He for three different periodicities of the modulated fibres, each 2.5 cm in length. Blue, 1 mm periodicity (Λ); red, 0.75 mm periodicity; green, 0.5 mm periodicity. All spectra were taken through two 0.2- μm zirconium filters, which rejected the laser light. The gas pressure was 111 torr, and the laser intensity was about $5 \times 10^{14} \text{ W cm}^{-2}$. The different curves were taken at different positions of the EUV grating in the EUV spectrometer, owing to the finite spectral region that can be captured simultaneously on the CCD. Therefore, the cut-off at low energy is artificial, limited by the spectral window. The high-energy cut-off is limited by the available laser intensity. It is expected to increase significantly when longer modulated sections, lower pressures, shorter laser pulses (10 fs), or higher laser intensities are used.

signal originates from a short window in time within the driving pulse. In QPM, this critical ionization level is higher, and occurs closer in time to the peak of the pulse if the optimal laser intensity is used. Here, the ionization rate is highest, resulting in a very fast 'sweep' of the phase-matching conditions through the optimum. Simple calculations show that this time window can be significantly shorter than for the equivalent straight fibre, and may limit the HHG emission to a single half-cycle of the driving laser. This phenomenon of time-gating of HHG by means of rapidly varying phase-matching conditions is new, and is distinct from the concept of time-gating using intensity variations in a short, sub-10-fs pulse^{22,23}.

Finally, by using either higher intensity laser pulses, longer modulated sections, lower pressures, or shorter (10 fs) pulses, our calculations show that it should be possible to extend this work to phase match HHG at wavelengths well below 4.4 nm. This wavelength is in the 'water window' region of the spectrum that is relevant to high-contrast biological imaging; HHG has been shown to reach these short wavelengths, but only under inefficient non-phase-matched conditions^{1,2}. We expect that this phase-matched EUV source will thus make possible element-specific spectroscopies and microscopies with ultrahigh spatial and temporal resolution. □

Received 23 July; accepted 15 October 2002; doi:10.1038/nature01222.

1. Chang, Z. H., Rundquist, A., Wang, H. W., Murnane, M. M. & Kapteyn, H. C. Generation of coherent soft X-rays at 2.7 nm using high harmonics. *Phys. Rev. Lett.* **79**, 2967–2970 (1997).
2. Spielmann, C. *et al.* Generation of coherent X-rays in the water window using 5-femtosecond laser pulses. *Science* **278**, 661–664 (1997).
3. McPherson, A. *et al.* Studies of multiphoton production of vacuum-ultraviolet radiation in the rare gases. *J. Opt. Soc. Am. B* **4**, 595–601 (1987).
4. Ferray, M. *et al.* Multiple-harmonic conversion of 1064 nm radiation in rare gases. *J. Phys. B* **21**, L31–L35 (1987).
5. Rundquist, A. *et al.* Phase-matched generation of coherent soft X-rays. *Science* **280**, 1412–1415 (1998).
6. Bloembergen, N. *Nonlinear Optics* (World Scientific, River Edge, NJ, 1996).
7. Franken, P. A., Hill, A. E., Peters, C. W. & Weinreich, G. Generation of optical harmonics. *Phys. Rev. Lett.* **7**, 118–119 (1961).
8. Miller, G. D. *et al.* 42%-efficient single-pass cw second-harmonic generation in periodically poled lithium niobate. *Opt. Lett.* **22**, 1834–1836 (1997).
9. Fejer, M. M., Magel, G. A., Jundt, D. H. & Byer, R. L. Quasi-phase-matched 2nd harmonic-generation — Tuning and tolerances. *IEEE J. Quant. Electron.* **28**, 2631–2654 (1992).
10. Zhu, S., Zhu, Y. Y. & Ming, N. B. Quasi-phase-matched third-harmonic generation in a quasi-periodic optical superlattice. *Science* **278**, 843–846 (1997).
11. Armstrong, J. A., Bloembergen, N., Ducuing, J. & Pershan, P. S. Interactions between light waves in a nonlinear dielectric. *Phys. Rev. A* **127**, 1918–1939 (1962).
12. Bartels, R. A. *et al.* Fully spatially coherent EUV beams generated using a small-scale laser. *Science* **297**, 376–378 (2002).
13. Bartels, R. *et al.* Shaped-pulse optimization of coherent soft X-rays. *Nature* **406**, 164–166 (2000).
14. Durfee, C. G. *et al.* Phase matching of high-order harmonics in hollow waveguides. *Phys. Rev. Lett.* **83**, 2187–2190 (1999).
15. Backus, S. *et al.* High-efficiency, single-stage 7-kHz high-average-power ultrafast laser system. *Opt. Lett.* **26**, 465–467 (2001).
16. Christov, I. P., Kapteyn, H. C. & Murnane, M. M. Quasi-phase matching of high-harmonics and attosecond pulses in modulated waveguides. *Opt. Express* **7**, 362–367 (2000).
17. Shkolnikov, P. L., Lago, A. & Kaplan, A. E. Optimal quasi-phase-matching for high-order harmonic generation in gases and plasmas. *Phys. Rev. A* **50**, R4461–R4464 (1994).
18. Lange, H. R. *et al.* High-order harmonic generation and quasiphasematching in xenon using self-guided femtosecond pulses. *Phys. Rev. Lett.* **81**, 1611–1613 (1998).
19. Misoguti, L. *et al.* Generation of broadband VUV light using third-order cascaded processes. *Phys. Rev. Lett.* **87**, 013601–013604 (2001).
20. Voronov, S. L. *et al.* Control of laser high-harmonic generation with counterpropagating light. *Phys. Rev. Lett.* **87**, 133902–133905 (2001).
21. Somekh, S. & Yariv, A. Phase matching by periodic modulation of the nonlinear optical properties. *Opt. Commun.* **6**, 301–304 (1972).
22. Christov, I. P., Murnane, M. M. & Kapteyn, H. C. High-harmonic generation of attosecond pulses in the "single-cycle" regime. *Phys. Rev. Lett.* **78**, 1251–1254 (1997).
23. Hentschel, M. *et al.* Attosecond metrology. *Nature* **414**, 509–513 (2001).

Acknowledgements This work was supported by the National Science Foundation and the Department of Energy.

Competing interests statement The authors declare competing financial interests: details accompany the paper on Nature's website (► <http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to M.M.M. (e-mail: murnane@jila.colorado.edu).

Electroluminescent device with reversible switching between red and green emission

S. Welter*, K. Brunner†, J. W. Hofstra†* & L. De Cola*

* Molecular Photonics Group, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

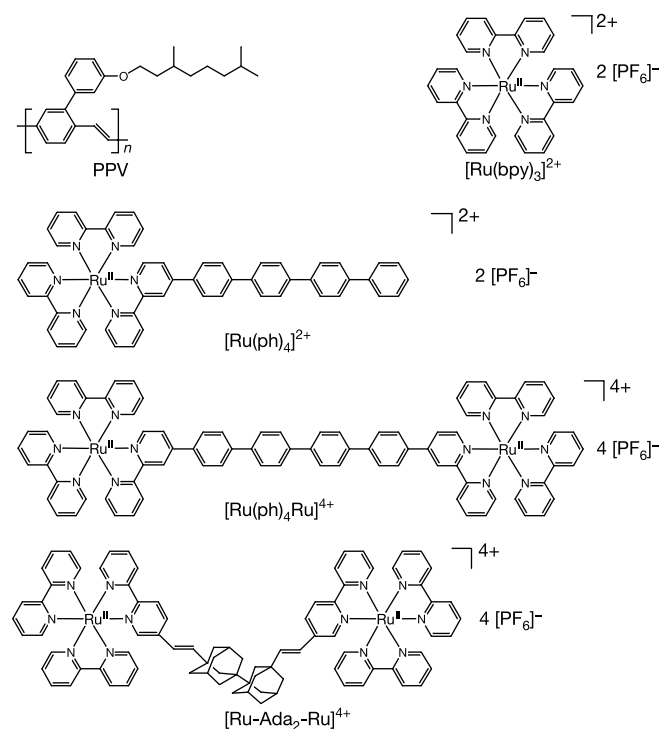
† Philips Natuurkundig Laboratorium, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

Research on new materials for organic electroluminescence has recently focused strongly on phosphorescent emitters^{1–3}, with the aim of increasing the emission efficiency and stability. Here we report the fabrication of a simple electroluminescent device, based on a semiconducting polymer combined with a phosphorescent complex, that shows fully reversible voltage-dependent switching between green and red light emission. The active material is made of a polyphenylenevinylene (PPV) derivative molecularly doped with a homogeneously dispersed dinuclear ruthenium complex, which fulfils the dual roles of triplet emitter and electron transfer mediator. At forward bias (+4 V), the excited state of the ruthenium compound is populated, and the characteristic red emission of the complex is observed. On reversing the bias (–4 V), the lowest excited singlet state of the polymer host is populated, with subsequent emission of green light. The mechanism for the formation of the excited state of the PPV derivative involves the ruthenium dinuclear complex in a stepwise electron transfer process that finally leads to efficient charge recombination reaction on the polymer.

The work of Bard and co-workers⁴ on the electroluminescence of ruthenium complexes in solution prepared the ground for the design and development of solid-state devices based on metal complexes. The Ru complexes form an interesting class of electroluminescent materials. Red emission, centred around 620 nm, high external efficiency^{5–7} at low voltages, easy synthesis and good processability are attractive properties of these complexes⁸, which make them eminent candidates for application as dopant in emissive polymer layers. The polypyridine Ru complexes are intrinsically charged. The mechanism associated with their electroluminescence is based on charge injection across the electrodes via an electrochemical redox pathway, and therefore is independent of the work function of the electrodes. Such devices are referred to as light-emitting electrochemical cells⁹.

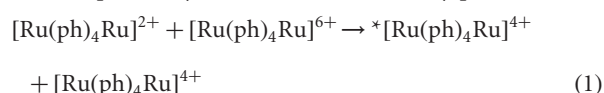
A dinuclear ruthenium complex, referred to here as [Ru(ph)₄-Ru]⁴⁺ (see below for full structure), containing two Ru(bpy)₃²⁺ moieties (where bpy is 2,2'-bipyridine) and four phenylene (ph) units as spacers between the two ruthenium centres, was mixed (Ru/PPV mass ratio 1:5) with a green-emitting polyphenylenevinylene derivative, poly(2-(*m*-3,7-dimethyloctyloxy-phenyl)-*p*-phenylenevinylene, referred to here as 'PPV'). The light-emitting polymer and the dinuclear ruthenium complex are applied such that they form a homogeneous, amorphous layer on a scale that can be observed by visual inspection and optical luminescence microscopy of the layer. For reference purposes and to assess the proposed mechanism (see below), three other complexes, referred to here as [Ru(bpy)₃]²⁺, [Ru(ph)₄]²⁺ and [Ru-Ada₂-Ru]⁴⁺ (ref. 10), were used (see below for full structures). No lateral phase separation is observed in the film formed by the polymer doped with the metal complexes (as measured, for example, by secondary ion mass spectrometry).

The highest occupied molecular orbital (HOMO) levels of the polymer and of the ruthenium compound are almost isoenergetic, whereas the lowest unoccupied molecular orbital (LUMO) of the Ru complex is typically about 0.5 eV lower in energy than that of the



polymer. The device emits pure red light from the Ru complex when the indium tin oxide (ITO) contact is biased positively (+4 V, forward bias), and green light from the polymer when the ITO contact is biased negatively (−4 V, reverse bias), as shown in Fig. 1. This layer stack proves to be intrinsically asymmetric, as demonstrated by the independence of the device from the mode of operation. In other words, the device always emits red light under forward bias and green under reverse bias, regardless of whether it has been cycled first to forward or reverse bias. Evidence that in our case we are not dealing with a conventional electrochemical cell is that addition of an inert, conducting salt to the emissive layer (comprising the PPV doped with the metal complex) changes the device properties considerably: emission of red light at forward and at reverse bias is observed (the effect of added inert salt is discussed later in more detail).

The reason for the absence of green polymer emission and the presence of only red luminescence from the ruthenium unit at forward bias lies in the mechanism of the population of the excited state of the emitting compound. Upon application of a low positive voltage (Fig. 2), hole injection and electron injection occur from the ITO and Au electrodes, respectively. At the ITO electrode, oxidation of ruthenium (from 2+ to 3+) and oxidation of PPV can be achieved, while at the Au electrode, electrochemical reduction of the ruthenium compound takes place. At the voltages applied (<5 V), it is impossible in undoped devices to inject electrons from Au contacts into the PPV LUMO level, and consequently no green emission from the polymer is observed. Ruthenium complexes—either as pure film or admixed in polymeric matrices—are known to emit via a well-understood electroluminescent reaction between the oxidized and the reduced ruthenium complexes^{2,11,12}. So the reaction that produces the excited state of the ruthenium complex in the present system can be schematically presented as:



The concentration gradient of Ru redox products is followed by a

redistribution of mobile counter ions (PF_6^-) to ensure overall charge compensation. At the boundary region of Ru^{3+} and Ru^{+} , charge recombination occurs between Ru^{3+} and Ru^{+} , and red light (luminance of 25 cd m^{-2}) is emitted¹¹.

Under reverse bias, the mechanism of light generation is completely different. Green light is observed already at low voltages (−4 V; Fig. 2) for $[\text{Ru}(\text{ph})_4\text{Ru}]^{4+}$ complexes and under similar conditions also for $[\text{Ru}(\text{ph})_4]^{2+}$ complexes. It is known that a pristine PPV device cannot give any emission under these conditions, as the electron injection barrier at the ITO interface is too high. Also, reduced PPV cannot be formed in the blend comprising the polymer and the Ru complex unless the $(\text{PF}_6)^-$ counter ions cause most of the field to drop at the electrode interface or dope the polymer and thereby facilitate charge injection, analogous to the conditions in a light-emitting electrochemical cell^{9,13}.

In order to establish whether this mechanism is active in our device, we examined the role of the $(\text{PF}_6)^-$ ions in the population of the polymer LUMO by substituting the dinuclear complex with the mononuclear complex $\text{Ru}(\text{bpy})_3^{2+}$. Great care was taken to always use the same concentration of chromophore (that is, Ru unit) when comparing mono- and dinuclear complexes, which also automatically results in the same concentration of counter ions. No light emission at reverse bias is observed at the voltages used (less than −9 V) when the PPV is doped with the $\text{Ru}(\text{bpy})_3^{2+}$ complex, and therefore a simple salt effect involving the counter ions in an ionic gradient at the electrode can be ruled out as a mechanism for the population of the polymer LUMO. On the other hand, if in addition to the Ru complex the inert salt $t\text{-Bu}_4\text{NPF}_6$ is added, efficient red emission is seen at forward and reverse bias. This means that the device switches upon addition of an inert salt to a conventional, symmetric, electrochemical cell, and the differences in the injection behaviour of the Au and ITO contacts are removed by the presence of the inert conducting salt (see also below for devices with symmetric contacts). The mechanism involved in the production of green light at reverse bias for the device containing the $[\text{Ru}(\text{ph})_4\text{Ru}]^{4+}$ compound is therefore not related to the standard behaviour of a PPV-based electrochemical cell. This is further supported by the fact that direct injection of negative charge into the PPV LUMO is unlikely, because the Ru complexes are reduced at lower potential than the polymer.

The polymer used in the present work was chosen—among other reasons—because of the absence of a considerable contribution of energy transfer, upon light excitation, from the polymer to the dinuclear and mononuclear Ru complexes that are applied as

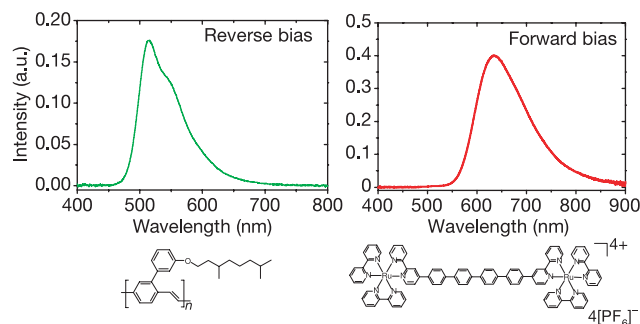


Figure 1 Electroluminescence spectra at forward bias (right) and reverse bias (left) of a light-emitting cell comprising a dinuclear Ru complex, $[\text{Ru}(\text{ph})_4\text{Ru}]^{4+}$, mixed in a PPV host matrix as the emissive layer. The green emission at reverse bias (−4 V) is due to charge recombination on the PPV polymer, whereas the red emission at forward bias (+4 V) results from the dinuclear Ru complex. Below the spectra, the structures of the emitting species are shown.

dopants. This is a prerequisite for green emission to occur at reverse bias. At the ITO electrode, the reduction of the Ru complex can occur with formation of mono- or bi-reduced species. Reduction of the Ru-based components is not centred on the metal, but on the ligands⁸. In particular, the first reduction of $[\text{Ru}(\text{ph})_4\text{Ru}]^{4+}$ occurs on the bipyridines linked to the polyphenylene units, and is bielectronic in nature. Upon reduction, the polyphenylene bridge can redistribute the charge in such a way that a stable bipolaron can be formed, and further stabilization is provided by the cation (Ru^{2+}) in its proximity (Fig. 3). The reduced species is stabilized by the presence of the polyphenylene units, as it is well-known that for such aromatic species, a bipolaron—either negative or positive—can be stabilized by the resonant quinoid structure¹⁴ that can be formed upon double reduction.

Clear experimental evidence for the nature of the species formed is obtained from the spectroelectrochemistry of the dinuclear ruthenium complex in butyronitrile solution. At -1.27 V versus the standard calomel electrode (SCE) the two bipyridines of the bridging ligand are reduced simultaneously (bielectronic wave), as shown by the decrease of the absorption band at 290 nm . Furthermore, the transition related to the polyphenylenes at 345 nm is also sensitive to the reduction, with a slight shift towards lower energy ($\Delta E = 0.1\text{ eV}$) and an increase in intensity (higher absorbance). The metal-to-ligand charge transfer ($^1\text{MLCT}$) band of the Ru complex at 460 nm shifts markedly to lower energy ($\Delta E = 0.26\text{ eV}$), suggesting a strong electronic interaction between the two Ru centres of the dinuclear complex. This means that a more delocalized structure of the bridging ligand is formed upon a two-electron reduction of the $[\text{Ru}(\text{ph})_4\text{Ru}]^{4+}$ complex (see Fig. 3 for the proposed chemical structure), as compared with the unreduced species.

The double-reduced species can therefore directly transfer an electron to the oxidized PPV to create the excited state of the polymer, or the bridged Ru complex can mediate a stepwise electron transfer from the electrode to the PPV to generate the reduced PPV⁻, that otherwise cannot be obtained. With either mechanism, green light is generated at reverse bias. So we suggest the following

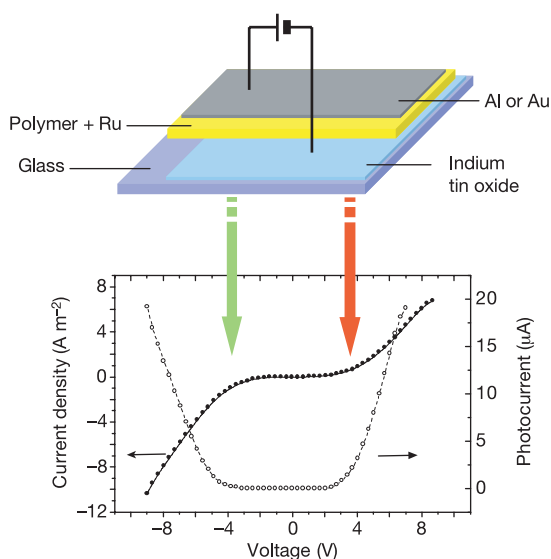


Figure 2 Schematic representation of the device structure (top), and plots of the current density (left) and photocurrent (right) versus voltage for the light-emitting cell described in Fig. 1. The homogeneous blend (Ru/PPV mass ratio 1:5) was spun from a dichloromethane solution on an ITO (indium tin oxide)-covered glass substrate, resulting in a thickness of 80 nm , and a metal contact was then evaporated on top of this layer stack. The device does not behave as a diode, but rather shows a nearly symmetric current versus voltage behaviour; it emits red light at forward bias, and green light at reverse bias.

mechanism for the green emission at reverse bias (Fig. 3):



The switching behaviour of the emission upon voltage change (red at $+4\text{ V}$ and green at -4 V ; compare Figs 1 and 2) is therefore related to the choice of the components of the blend and the hole-injecting electrode. As the ph_4 bridging unit is the crucial structural element, it is also reasonable to think that a mononuclear compound with a bridge attached (that can stabilize the additional charge) could lead to the same observation. We therefore tested the compound with four phenylene units attached, $[\text{Ru}(\text{ph})_4]^{2+}$, and indeed found that it gave green emission at reverse bias, although at lower intensities than the dinuclear species.

The mechanism is further confirmed by experiments using a dinuclear Ru complex similar in size and photophysical properties, but containing a saturated bridging ligand, $[\text{Ru-Ada}_2\text{-Ru}]^{4+}$ (see above). In such a case, as well as for the $[\text{Ru}(\text{bpy})_3]^{2+}$, no emission at reverse bias was observed when using a Ba/Al cathode. When Al or Au cathodes were used with the same device architecture and under the same conditions, we observed pure green emission (luminance of 330 cd m^{-2}) for the $[\text{Ru}(\text{ph}_4)\text{Ru}]^{4+}$ complex at reverse bias, whereas for the $[\text{Ru-Ada}_2\text{-Ru}]^{4+}$ complex, predominantly red emission is obtained.

Such behaviour cannot be attributed to different miscibilities of the various Ru complexes or to different microstructures of the films, as previously reported for other voltage-responsive polymer layers^{15–18}, because the compounds have the same charge and solubility, and similar chemical behaviour. But when the phase

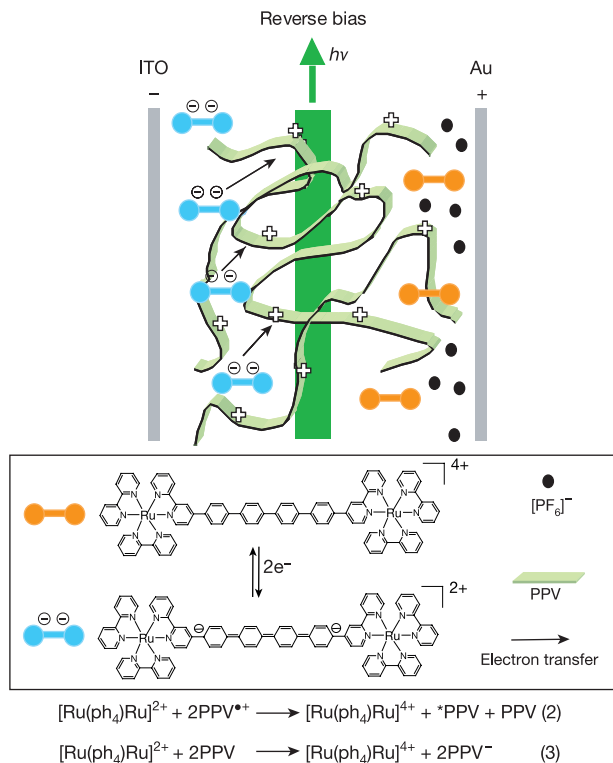


Figure 3 Proposed mechanism for the emission of green light at reverse bias. The doubly reduced Ru complex acts as electron transfer mediator to the PPV polymer. Green light is emitted when positive and negative charge carriers recombine radiatively on the PPV polymer. Blue, the reduced Ru complex; orange, the oxidized Ru complex; and green, PPV. Upon two-electron reduction of the $[\text{Ru}(\text{ph}_4)\text{Ru}]^{4+}$ complex, a bipolaron structure could result. Delocalization of the electrons on the bridging ligand due to the stable quinoid structure is deduced from spectro-electrochemistry data (see text for details).

behaviour of the various complexes is the same, but they exhibit profoundly different emission properties, we conclude that the specific nature of the bridge is the key to the colour switching phenomenon. It is well known that electron conduction, stabilization and delocalization on such bridging ligands depend on the geometry, energetics and structural features of the molecules^{19–21}. Furthermore, use of different polymers with a higher-energy LUMO level (the HOMO level is chosen to be the same)—so that electron transfer mediated by the Ru complex cannot populate their excited state—did not result in green emission at reverse bias.

Symmetric devices with Au as anode and cathode have symmetric emission properties (that is, they show red emission at both forward and reverse bias), clearly indicating that an asymmetry in the device's charge injection behaviour is needed for the differentiation of the two possible mechanisms of light emission in the [Ru(phen)₄-Ru]⁴⁺ PPV system. As the work functions of the ITO and Au contacts are energetically comparable, and the red/green switching behaviour is also seen when Au is replaced by Al, influences other than simple energetics must also have a role in allowing the stepwise electron transfer at reverse bias; we intend to study these influences in greater detail in the near future. □

Received 19 July; accepted 12 November 2002; doi:10.1038/nature01309.

- Baldo, M. A. *et al.* Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* **395**, 151–154 (1998).
- Buda, M., Kalyuzhny, G. & Bard, A. J. Thin-film solid-state electroluminescent devices based on tris(2,2'-bipyridine)ruthenium(II) complexes. *J. Am. Chem. Soc.* **124**, 6090–6098 (2002).
- Wu, A., Yoo, D., Lee, J.-K. & Rubner, M. F. Solid-state light emitting devices based on the tris-chelated ruthenium(II) complex. 3. High efficiency devices via a layer-by-layer molecular-level blending approach. *J. Am. Chem. Soc.* **121**, 4883–4891 (1999).
- Faulkner, L. R. & Bard, A. J. *Electroanalytical Chemistry* (ed. Bard, A. J.) 1–95 (Marcel Dekker, New York, 1977).
- Rudmann, H., Shimada, S. & Rubner, M. F. Solid-state light emitting devices based on the tris-chelated ruthenium(II) complex. 4. High-efficiency light-emitting devices based on derivatives of the tris(2,2'-bipyridyl) ruthenium(II) complex. *J. Am. Chem. Soc.* **124**, 4918–4921 (2002).
- Handy, E. S., Pal, A. J. & Rubner, M. F. Solid-state light emitting devices based on the tris-chelated ruthenium(II) complex. 2. Tris(bipyridyl)ruthenium(II) as a high-brightness emitter. *J. Am. Chem. Soc.* **121**, 3525–3528 (1999).
- Elliott, C. M., Pichot, F., Bloom, C. J. & Rider, L. S. Highly efficient solid-state electrochemically generated chemiluminescence from ester-substituted trisbipyridineruthenium(II)-based polymers. *J. Am. Chem. Soc.* **120**, 6781–6784 (1998).
- Juris, A. *et al.* Ru(II) polypyridine complexes: photophysics, photochemistry, and chemiluminescence. *Coord. Chem. Rev.* **84**, 85–277 (1988).
- Pei, Q., Yu, G., Yang, Y. & Heeger, A. J. Polymer light-emitting electrochemical cells. *Science* **269**, 1086–1088 (1995).
- De Cola, L. & Belser, P. Photoinduced energy and electron transfer processes in rigidly bridged dinuclear Ru/Os complexes. *Coord. Chem. Rev.* **177**, 301–346 (1998).
- Maness, K. M., Terrill, R. H., Meyer, T. J., Murray, R. W. & Wightman, R. M. Solid-state diode-like chemiluminescence based on serial, immobilized concentration gradients in mixed-valent poly[Ru(vbpy)₃](PF₆)₂ films. *J. Am. Chem. Soc.* **118**, 10609–10616 (1996).
- Luttmer, J. D. & Bard, A. J. Electrogenerated chemiluminescence. 38. Emission intensity-time transients in the tris(2,2'-bipyridine)ruthenium(II) system. *J. Phys. Chem.* **85**, 1155–1159 (1981).
- deMello, J. C., Tessler, N., Graham, S. C. & Friend, R. H. Ionic space-charge effects in polymer light-emitting diodes. *Phys. Rev. B* **57**, 12951–12963 (1998).
- Brédas, J. L., Chance, R. R. & Silbey, R. Comparative theoretical study of the doping of conjugated polymers: polarons in polyacetylene and polyparaphenylene. *Phys. Rev. B* **26**, 5843–5854 (1982).
- Berggren, M. *et al.* Light-emitting diodes with variable colours from polymer blends. *Nature* **372**, 444–446 (1994).
- Yang, Y. & Pei, Q. Voltage controlled two colour light-emitting electrochemical cells. *Appl. Phys. Lett.* **68**, 2708–2710 (1996).
- Hamaguchi, M. & Yoshino, K. Color-variable electroluminescence from multilayer polymer films. *Appl. Phys. Lett.* **69**, 143–145 (1996).
- Tasch, S. & Brandstätter, C. Red-green-blue light emission from a thin film electroluminescence device based on parahexaphenyl. *Adv. Mater.* **9**, 33–36 (1997).
- DeCola, L. & Belser, P. *Electron Transfer in Chemistry* (ed. Balzani, V.) 97–136 (Wiley-VCH, Weinheim, 2001).
- Yaliraki, S. N., Kemp, M. & Ratner, M. A. Conductance of molecular wires: influence of molecule-electrode binding. *J. Am. Chem. Soc.* **121**, 3428–3434 (1999).
- Pourtois, G., Beljonne, D., Cornil, J., Ratner, M. A. & Brédas, J. L. Photoinduced electron-transfer processes along molecular wires based on phenylenevinylene oligomers: a quantum-chemical insight. *J. Am. Chem. Soc.* **124**, 4436–4447 (2002).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.D.C. (e-mail: ldc@science.uva.nl) or K.B. (e-mail: klemens.brunner@philips.com).

Fingerprints of global warming on wild animals and plants

Terry L. Root*, Jeff T. Price†, Kimberly R. Hall‡, Stephen H. Schneider§, Cynthia Rosenzweig|| & J. Alan Pounds¶

* Center for Environmental Science and Policy, Institute for International Studies, Stanford University, Stanford, California 94305, USA

† American Bird Conservancy, 6525 Gunpark Drive, Suite 150-146, Boulder, Colorado 80301, USA

‡ Department of Fisheries and Wildlife, 13 Natural Resources Building, Michigan State University, East Lansing, Michigan 48824-1222, USA

§ Department of Biological Sciences & Institute for International Studies, Stanford University, Stanford, California 94305, USA

|| National Aeronautics and Space Administration, Goddard Institute for Space Studies, 2880 Broadway, Suite 750, New York, New York 10025, USA

¶ Golden Toad Laboratory for Conservation, Monteverde Cloud Forest Preserve and Tropical Science Center, Santa Elena, Puntarenas 5655, Box 73, Costa Rica

Over the past 100 years, the global average temperature has increased by approximately 0.6 °C and is projected to continue to rise at a rapid rate¹. Although species have responded to climatic changes throughout their evolutionary history², a primary concern for wild species and their ecosystems is this rapid rate of change³. We gathered information on species and global warming from 143 studies for our meta-analyses. These analyses reveal a consistent temperature-related shift, or 'fingerprint', in species ranging from molluscs to mammals and from grasses to trees. Indeed, more than 80% of the species that show changes are shifting in the direction expected on the basis of known physiological constraints of species. Consequently, the balance of evidence from these studies strongly suggests that a significant impact of global warming is already discernible in animal and plant populations. The synergism of rapid temperature rise and other stresses, in particular habitat destruction, could easily disrupt the connectedness among species and lead to a reformulation of species communities, reflecting differential changes in species, and to numerous extirpations and possibly extinctions.

Many studies have examined biological changes in relation to climatic change^{4,5}, but generally they are concentrated in particular regions or examine a limited set of taxa. To test whether or not a coherent pattern exists across regions and taxa that is consistent with predictions of expected change, we used two types of meta-analyses on these studies: vote counting and the regression-slope model (see Methods). One advantage of meta-analyses is that a broad spectrum of findings can be combined, including those for which statistical significance has not been shown. We examined thousands of articles, including those assembled by Working Group II of the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC TAR WGII)⁶. Of these, we included in our analyses only those that (1) examined a span of at least 10 years, (2) found that a trait of at least one species shows change over time, and (3) found either a temporal change in temperature at the study site or a strong association between the species trait and site-specific temperature. Because we were looking for trends, we also excluded studies examining climatic cycles, such as North Atlantic Oscillation and El Niño/Southern Oscillation. We divided 143 studies that met our criteria into two 'tiers': those demonstrating a statistically significant trend for at least one species examined (tier 1, most of which were used as the methodological basis for conclusions in the IPCC TAR WGII^{6,7}) and those in which statistical significance was not shown by the study's authors (tier 2), usually because no statistical tests were applied. We performed our analyses for each tier taken separately and for the two combined. Appendices 1 and 2 of the Supplementary Information provide the data and citations

Table 1 Summary statistics for meta-analyses

	Significant species	Nonsignificant species	Combined species
Number of species changing	586	882	1,468
Number changing in expected direction	482	708	1,190
Percentage in expected direction	82.3%	80.4%	81.1%
90% confidence interval for percentage in expected direction	73.4–88.6%	70.5–87.4%	74.2–86.5%
Effect size (δ)	–0.09	–0.23	–0.23
90% confidence interval for δ	–0.12 to –0.06	–0.30 to –0.14	–0.29 to –0.17
Standard error for δ	0.0004	0.0023	0.0014
Correlation coefficient (r)	–0.05	–0.12	–0.12
90% confidence interval for r	–0.06 to –0.03	–0.16 to –0.07	–0.15 to –0.09
Standard error for r	0.0002	0.0012	0.0007

A breakdown of values for those species or groups of species that were found, in the studies examined, to have statistically significant trends for various traits and for those that were not statistically significant. In addition, values are listed for the combination of these two categories of species or species groups.

for all of the studies used in our meta-analyses.

We focused on temperature change and ignored other climatic changes, such as precipitation, because the biological effects of temperature are often better understood for most of the organisms examined. Explicitly considering drought in our analyses would have allowed us to include many more studies, particularly from the Southern Hemisphere, but attributing local droughts to globally coherent patterns of climatic changes can often be difficult. We do, however, recognize that factors influencing populations interact in complex ways: temperature can exert its influence, for example, by affecting moisture availability.

Four types of change in species' traits due to warming may be possible. First, the density of species may change at given locations, and the ranges of species may shift either poleward or up in elevation as species move to occupy areas within their metabolic temperature tolerances. Second, because many natural history traits of species are triggered by temperature-related cues, changes could occur in the timing of events (phenology), such as migration, flowering or egg laying. Third, changes in morphology, such as body size, and behaviour may occur. Fourth, genetic frequencies may shift.

Attributing observed changes in populations of plants and animals to climatic change, specifically temperature increases, is possible because we expect the trends created by the large-scale pressure of global warming to show widespread, predictable and concordant patterns of change. In addition, we expect these changes to be concentrated in areas where temperature changes are largest (that is, at higher latitudes and altitudes) and for changes to be less evident elsewhere. Climate change is only one of a long list of pressures that influence the distributions and health of populations, as well as traits such as timing of activities and processes. These other pressures (for example, habitat modification, pollinator loss and exotic species introductions) often result in localized (often around centres of human populations) or multidirectional patterns of alterations to populations of species. To document a strong role for climate change in explaining many of the observed changes in

animal and plant populations, we looked for repeated examples occurring over long temporal and broad spatial scales that showed unidirectional changes predicted by our understanding of the physiological tolerances of species to temperature. The predicted result, or fingerprint, of an underlying consistent shift in a large-scale pattern shown by many species around the globe, coupled with an understanding of the possible causal mechanisms, provides confidence in attributing observed species changes to climatic change. The 85 tier 1 studies and 58 tier 2 studies found strong changes occurring around the globe in various types (taxa) of animals and plants (see Supplementary Information). The meta-analyses we used to analyse the findings of these studies were the vote-counting method⁸ and the regression-slope model⁹. We applied the vote-counting method to three different categories of data: 1) the 587+ species or groups of species ('+' because some studies do not provide numbers) in tier 1 (see Methods) that show statistically significant change; 2) the 886+ species or groups of species in both tier 1 and tier 2 that show statistically nonsignificant change or where the significance was unknown; and 3) the combined species (1,473+) showing change. For all three categories, the percentage of species changing in the expected direction was around 81% with a 90% confidence interval ranging from 71% to 89% (Table 1). The meta-analysis results of effect sizes and correlation coefficients were statistically different from zero ($P < 0.05$), which means that we can reject the null hypothesis of equal change in both the expected and opposite directions. Consequently, even this fairly low-powered vote-counting meta-analysis indicates that most changes are consistent with our understanding of how temperature change influences various traits of a variety of species and populations from around the globe. Hence, we can safely state that there has probably been a discernible impact of recent global warming on animals and plants.

The actual amount of change shown can be determined for those studies that examined shifts in spring phenologies. The 61 studies that investigated the change in timing of events occurring within the

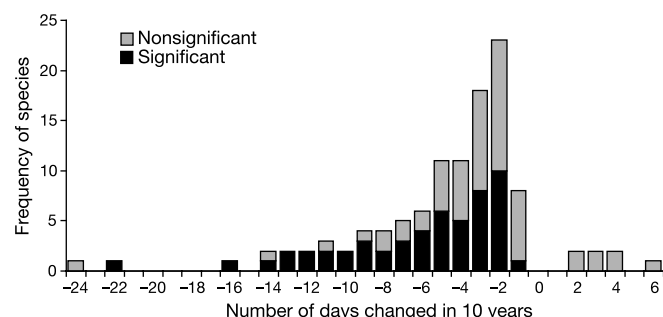


Figure 1 Frequency distribution of species and groups of species (see text) with a temperature-related trait changing by number of days in 10 years. No data were tabulated for species showing zero days changing in ten years (see Methods).

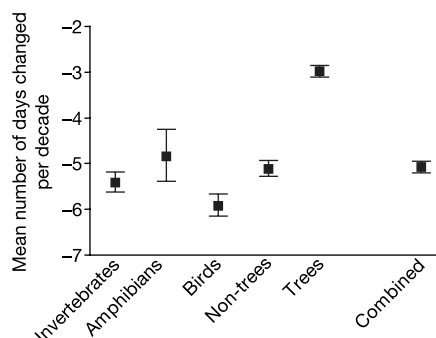


Figure 2 Means $\pm$ s.e.m. of days changed for the given groups of species. The 'Combined' category includes only those species tallied in the groups of species (that is, data for the one mammal, two fish and zooplankton are not included).

past 50 years examined a total of 694 species or groups of species. Our meta-analyses of these species indicate that over an average decade within the past 50 years, a statistically significant change towards earlier timing of spring events has occurred. The number of days changed per decade for a given species or species group ranges from 24 days earlier per decade for the breeding of North American common murre (*Uria aalge*) to 6.3 days per decade later for the breeding of North American Fowler's toad (*Bufo fowleri*) (Fig. 1). Using the regression-slope model⁹ (see Methods), we found that the estimated mean number of days changed per decade for all species showing change in spring phenology is 5.1 days earlier (s.e.m. ± 0.1 ; Fig. 1). When the data are grouped by statistical significance, the estimated mean of the species with nonsignificant findings is closer to zero than that for species with significant findings (-3.4 ± 0.1 ; -6.9 ± 0.1 , respectively; negative numbers indicate an earlier shift). The latter mean is understandably earlier than the former, given that regression analysis is more powerful at discriminating steeper slopes than shallower ones. However, all three estimated averages are statistically significantly different from zero, which means that these species are all showing a marked shift towards earlier spring events.

Given that higher latitudes have warmed more than the lower latitudes in the past half century (see Fig. 3d of ref. 1), we expect phenological responses to be larger nearer the poles and not as pronounced closer to the equator. (Unfortunately, our pool of literature did not allow us to test elevational changes.) The latitudes of the spring phenology studies in the Northern Hemisphere extend from 32° N to 72° N, with one study in the Southern Hemisphere at 39°. Because of the unevenness of the location of the data (for example, a preponderance of studies in the United Kingdom), we were able to have large enough sample sizes in each group only by dividing them into two groups along 50° latitude circles. The sample size from 32° to 49.9° was 24 + , which includes the one Southern Hemisphere study. From 50° to 72° N latitude, 85+ species or species groups were examined. As expected, the estimated mean and s.e.m. of the phenological shifts (see Methods) from 32° to 49.9° latitude is smaller (-4.2 ± 0.2) than that between 50° and 72° N latitude band (-5.5 ± 0.1). These two means are statistically significantly different from each other (Kruskal–Wallis test, ($P < 0.0001$), which strongly suggests that species at higher latitudes are indeed reacting more strongly to the more intense change in temperature.

Our spring phenology data set consists of species and populations from major taxa from molluscs to mammals. We had large enough sample sizes to examine the estimated means of the phenological shifts separately for invertebrates, amphibians and birds, and for trees and other plants (Fig. 2). Four of the five means cluster around an earlier shift of 5 days, which is the estimated mean for all taxa combined. Trees, however, show an estimated mean that is later than the cluster (-3.0 ± 0.1). This estimated mean is statistically different from the other means (Kruskal–Wallis test), and the other four means are not statistically different from one another.

Our study shows that recent temperature change has apparently already had a marked influence on many species. Meta-analyses provide a way to combine results, whether significant or not, from various studies, and to find an underlying consistent shift, or fingerprint, among species from different taxa examined at disparate locations. The findings for the nonsignificant species, when aggregated, show nearly as much significant change as the group of species showing statistical significance (Table 1 and Fig. 1). Hence, for the studies we examined, the balance of evidence suggests that a significant impact of recent climatic warming is discernible in the form of long-term, large-scale alterations of animal and plant populations. For example, the average shift in spring phenology (timing) of events, such as breeding or blooming, for temperate-zone species is 5.1 ± 0.1 days earlier in a decade. The observed consistent broad-scale patterns of changes in the expected direc-

tions (80% of species showing change) strongly suggest that recent temperature trends are the most likely explanation for these observed phenomena. Clearly, if such climatic and ecological changes are now being detected when the globe has warmed by an estimated average of only 0.6 °C, many more far-reaching effects on species and ecosystems will probably occur in response to changes in temperature to levels predicted by IPCC¹, which run as high as 6 °C by 2100.

Projected future rapid climate change could soon become a more looming concern, especially when occurring together with other already well-established stressors, particularly habitat destruction. During rapid climatic changes in the past, species showed differential movements¹⁰, rather than shifting together as suggested by many authors, including Darwin¹¹. Such differential movement could result in a disruption of the connectedness among many species in current ecosystems (for example, a tearing apart of communities⁷). Research and conservation attention needs to be focused not only on global warming and each of the other stressors by themselves, but also on the synergism of several pressures that together are likely to prove to be the greatest challenge to animal and plant conservation in the twenty-first century^{3,12}. Because anticipation of changes improves the capacity to manage—by acting proactively rather than reactively—it behoves us to increase our understanding about the responses of plants and animals to a changing climate. This understanding, coupled with further documentation of change, may well indicate a need for actions to modify conservation efforts and future planning to account for climate change, and to slow the projected rate of warming. □

Methods

We used results from 143 studies, each of which found some trait of a species showing a trend over a span of at least 10 years. The geometric mean of the time span for all studies was 30.3 years and the average was 34.5 years. Gaps between years were allowed. Several studies investigated more than one species and some species showed no change. The relative number of species exhibiting change compared with all species reported in the literature sources we examined is not the relevant metric we consider, because we are not trying to determine what percentage of species is responding to current climatic changes. Given that all species examined in this study (and indeed by the entire scientific community) represents only a small proportion of the total number of species that exist in the world (itself unknown), such percentage claims are untenable by any analysis. Rather, a relevant metric to detect a discernible influence of global warming on plants and animals is the fraction of those species exhibiting change that have changed in the direction expected given a temperature trend at their location. For those studies finding change in more than one species and reporting the change as an average for several species, only one entry is used in the various tests performed. For example, Fleming and Tatchell¹³ reported the change in flight period of five aphids as 2.6 days earlier per decade. To be conservative, for all our analyses, this group of aphids (and all similarly grouped species) is considered as one entry, rather than as five.

Meta-analyses used

We used two types of meta-analysis, vote counting and regression slope, to determine whether changes observed are consistent with the possibility that one force, global warming, is instigating a noticeable change in species. The vote-counting method is explained in detail in ref. 8. This method is biased towards finding zero or no effect¹⁴—of global warming, in our case. The various studies have different sampling lengths (K_i) and so we use a geometric mean of these numbers to determine K , the mean length of years the studies examined traits of species or populations. In all three vote-counting analyses that we used—species showing statistically significant change, species showing nonsignificant change, and the combined species that showed change—the value of K is 30.

The regression-slope meta-analysis provides a way to examine directly the shift in spring phenological changes. For this method to work, the various studies examined must have used the same measurement of change, which in these meta-analyses is the number of days changed per decade. We included only those studies that examined more recent spring shifts, from 1951 to 2001. Details of the methods we used here to derive estimated slopes are given in ref. 9. Our only deviation from this formulation is that we did not include the sampling variance term in the calculation of the variance of the slope parameter. This is because most studies examined do not report some of the data necessary to derive the sampling variance. This term acts to reduce the size of the variance of the regression slope. Consequently, the variances presented here are slight overestimates, making our inferences more conservative.

Potential biases

Apart from the problems inherent in summarizing information from diverse studies of numerous subjects and methods, potential biases in the data are also of concern in analyses like ours. We do not claim that all authors of studies in the literature we cite report all

species they observe, and there may be a bias to report primarily those species that show change. Even if there were such a bias, however, it would have no influence on our claim of a discernible impact of warming on plants and animals, because our metric of investigation is what fraction of those species that exhibit change has changed in the direction expected with local temperature trends, not what fraction of all species has exhibited change. The only way that observer bias could influence our metric would be if there were a systematic bias among the scores of studies we examine for researchers to select as study subjects only species showing changes in the direction preconceived by the authors to reflect temperature change. In addition, these many authors would have to have deliberately and systematically suppressed reporting on those species that changed in directions opposite to that expected. We find this possibility of widespread and systematic biases far-fetched, and thus believe that the metric we use is adequate for examining in an unbiased manner the existence of a discernible climatic signal in the traits of many plants and animals.

Received 12 September; accepted 26 November 2002; doi:10.1038/nature01333.

- Houghton, J. T. et al. (eds) in *Climate Change 2001: The Science of Climate Change* (Cambridge Univ. Press, New York, 2001).
- Harris, A. H. Wisconsinian pre-pleniglacial biotic change in southeastern New Mexico. *Quat. Res.* **40**, 127–133 (1993).
- Schneider, S. H. & Root, T. L. in *Status and Trends of the Nation's Biological Resources Part 1* (eds Mac, M. J., Opler, P. A., Puckett Haacker, N. & Doran, P. D.) 89–116 (US Geological Survey, Reston, Virginia, 1998).
- Walther, G.-R. et al. Ecological responses to recent climate change. *Nature* **416**, 389–395 (2002).
- Penuelas, J. & Filella, I. Responses to a warming world. *Science* **294**, 793–795 (2001).
- McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J. & White, K. S. (eds) *Climate Change 2001: Impacts, Adaptations, and Vulnerability* (Cambridge Univ. Press, New York, 2001).
- Root, T. L. & Schneider, S. H. in *Wildlife Responses to Climate Change: North American Case Studies* (eds Schneider, S. H. & Root, T. L.) 1–56 (Island Press, Washington DC, 2002).
- Hunter, J. E. & Schmidt, F. L. *Methods of Meta-Analysis: Correcting Errors and Bias in Research Findings* (Sage, Thousand Oaks, California, 1995).
- Raju, N. S., Fralich, R. & Steinhaus, S. D. Covariance and regression slope models for studying validity generalization. *Appl. Psychol. Meas.* **10**, 195–211 (1986).
- Overpeck, J. T., Webb, R. S. & Webb, T. III Mapping eastern North American vegetation change over the past 18,000 years: no analogs and the future. *Geology* **20**, 1071–1074 (1992).
- Darwin, C. *On the Origin of Species: A Facsimile of the First Edition* (Harvard Univ. Press, Cambridge, Massachusetts, 1964).
- Myers, N. in *Biodiversity II: Understanding and Protecting our Biological Resources* (eds Reaka-Kudla, M. L., Wilson, D. E. & Wilson, E. O.) 125–138 (Joseph Henry, Washington DC, 1997).
- Fleming, R. A. & Tatchell, G. M. in *Insects in a Changing Environment* (eds Harrington, R. & Stork, N. E.) 505–508 (Academic, San Diego, California, 1995).
- Hedges, L. V. & Olkin, I. *Statistical Methods for Meta-Analysis* (Academic, New York, 1985).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements T.L.R. acknowledges partial support from the US Environmental Protection Agency, the Winslow Foundation, the University of Michigan's Office of the Vice President for Research, and the School of Natural Resources and Environment. S.H.S. acknowledges partial support from the Winslow Foundation. We thank T. Haff, J. Jeffries, M. Kirpes, J. Manternach, T. Reed, A. Stover, C. Wood and J. Cary for their efforts. In addition, we thank M. Apps, J. Magnuson and C. Parmesan for stimulating discussions and for help in finding obscure articles.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T. R.
(e-mail: troot@stanford.edu).

True navigation and magnetic maps in spiny lobsters

Larry C. Boles & Kenneth J. Lohmann

Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA

Animals are capable of true navigation if, after displacement to a location where they have never been, they can determine their position relative to a goal without relying on familiar surroundings, cues that emanate from the destination, or information collected during the outward journey^{1,2}. So far, only a few animals, all vertebrates, have been shown to possess true navigation³. Those few invertebrates that have been carefully studied

return to target areas using path integration, landmark recognition, compass orientation and other mechanisms that cannot compensate for displacements into unfamiliar territory^{4,5}. Here we report, however, that the spiny lobster *Panulirus argus* oriented reliably towards a capture site when displaced 12–37 km to unfamiliar locations, even when deprived of all known orientation cues *en route*. Little is known about how lobsters and other animals determine position during true navigation. To test the hypothesis that lobsters derive positional information from the Earth's magnetic field, lobsters were exposed to fields replicating those that exist at specific locations in their environment. Lobsters tested in a field north of the capture site oriented themselves southwards, whereas those tested in a field south of the capture site oriented themselves northwards. These results imply that true navigation in spiny lobsters, and perhaps in other animals, is based on a magnetic map sense.

In the context of homing behaviour, an animal capable of true navigation must possess both a positional sense to determine its location and a directional or compass sense to orient in the appropriate homeward direction^{6,7}. Many animals, both vertebrate and invertebrate, possess diverse compasses based on the Earth's magnetic field, the position of the Sun, patterns of skylight polarization and the positions of stars^{4,8}. In contrast, few animals are known to possess the ability to determine position relative to a goal after being displaced to unfamiliar areas under conditions in which

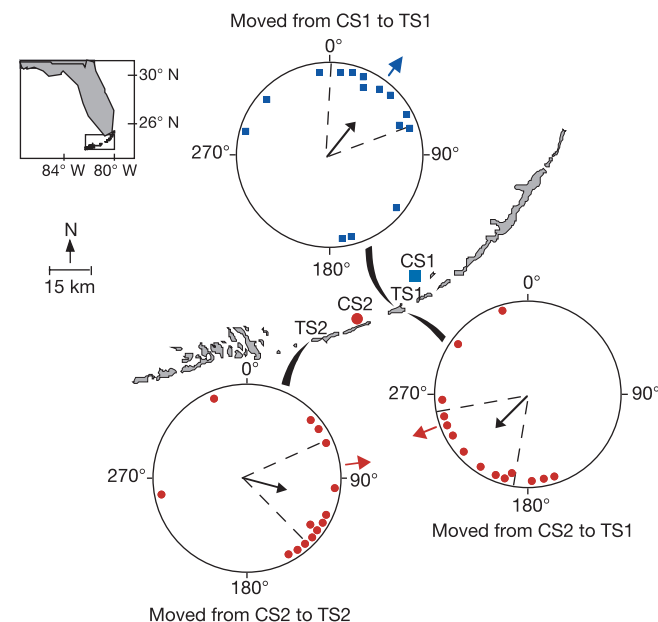


Figure 1 Orientation of displaced lobsters. Lobsters were transported by boat from two capture sites (CS1, CS2) via circuitous routes (see the text) to one of two test sites (TS1, TS2). In the orientation diagrams, each small symbol represents the mean angle of a single lobster. Blue squares indicate lobsters captured at CS1, whereas red circles indicate lobsters captured at CS2. The arrow in the centre of each orientation diagram indicates the mean angle of each group; the arrow length is proportional to the mean vector r , with the radius of the circle corresponding to $r = 1$. Lobsters transported from CS1 to TS1 were significantly oriented ($r = 0.51$, $Z = 3.92$, $P < 0.02$, Rayleigh test) with a mean angle of 38° . Lobsters transported from CS2 to TS1 were significantly oriented ($r = 0.65$, $Z = 5.96$, $P < 0.01$) with a mean angle of 222° . Lobsters displaced from CS2 to TS2 were also significantly oriented ($r = 0.51$, $Z = 3.89$, $P < 0.02$) with a mean angle of 105° . In all orientation diagrams, the dashed lines represent the 95% confidence interval for the mean angle. Data are plotted relative to magnetic north. The blue or red arrow outside each orientation diagram indicates the direction from the test site to the capture site. In each case, the mean angle of orientation coincided closely with the direction towards the capture site (see the text) and the 95% confidence interval encompassed this 'homeward' direction.

they cannot monitor the outward journey to the test site. Until now, evidence for true navigation has been limited to birds and a few other specialized migrants, all of them vertebrates³.

Little is known about the sensory cue or cues that underlie true navigation. The navigational system of homing pigeons has been studied for decades, yet the nature of their position-finding system remains the subject of a long-standing and unresolved debate^{9,10}. In principle, animals might derive positional information from one or more elements of the Earth's magnetic field^{2,11}. However, little direct evidence exists that true navigation is based on magnetic cues, and whether magnetic maps exist at all has remained controversial^{12,13}.

The Caribbean spiny lobster, *Panulirus argus*, is a migratory crustacean that is capable of homing¹⁴. Juvenile and adult lobsters spend daylight hours hidden inside coral reef crevices or holes, emerging at night to forage over a considerable area before returning in darkness to the same den or to one of several others nearby¹⁵. Tag-recapture studies have indicated that lobsters can return to a home area after being displaced to unfamiliar locations¹⁶. These considerations led us to investigate whether spiny lobsters are capable of true navigation.

In an initial series of experiments, juvenile lobsters in Florida Bay, USA, were captured by divers and placed into opaque plastic containers partly filled with sea water. The containers were covered so that animals inside were deprived of visual cues. Lobsters were then transported by boat for 45–60 min along circuitous routes of up to 30 km, ending at one of two testing sites (Fig. 1). On the following morning, the eyestalks of the lobsters were covered with rubber caps to deprive them of visual cues during testing¹⁷, and each animal was tethered to a tracking system in the centre of a circular water-filled arena. The tracking system monitored the direction that each lobster walked.

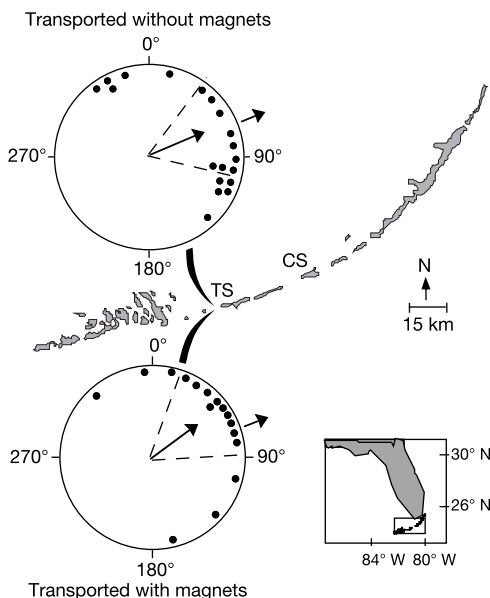


Figure 2 Orientation of lobsters transported overland by lorry in distorted magnetic fields. All lobsters in this experiment were collected from a single capture site (CS) near Long Key and transported to the test site (TS) at Pigeon Key. One group of lobsters (upper diagram) was subjected only to the magnetic field distortions produced by the metal body of the lorry. This group was significantly oriented ($r = 0.62$, $Z = 7.31$, $P < 0.001$, Rayleigh test) with a mean angle of 68° . The second group (lower diagram) was subjected to stronger, additional field distortions caused by stationary and moving magnets (see the text). This group was significantly oriented ($r = 0.67$, $Z = 7.31$, $P < 0.001$, Rayleigh test) with a mean angle of 53° . The arrows outside each orientation diagram indicate the direction from the test site to the capture site. Statistical conventions are as in Fig. 1. The two distributions were not significantly different ($U^2 = 0.143$, Watson test²⁹); instead, both groups of lobsters oriented themselves approximately towards the capture site (71°).

Lobsters captured at a location (CS1 on Fig. 1) 12 km to the north/northeast of the first test site (TS1 on Fig. 1) were significantly oriented with a mean angle of 38° . This direction corresponded closely to the most direct route (36°) back to the area of capture (Fig. 1). Lobsters captured at a site (CS2 on Fig. 1) 14 km west-southwest of the first testing site also were significantly oriented but with a mean angle of 222° ; this direction also corresponded closely to the most direct path back (250°). An additional group of lobsters captured at this same location was transported to a second testing site 18 km away (TS2 on Fig. 1), where the most direct heading to the capture area was 82° . These lobsters were also significantly oriented with a mean angle of 105° . Thus, in each case, lobsters with their eyes covered walked in directions that would have led them back towards the area where they were captured.

The definition of true navigation requires that animals derive positional information from local cues available at the test site rather than by exploiting information gathered during the outward journey^{2,3}. In our initial experiments, lobsters oriented towards the capture site despite being transported in sealed containers under conditions in which no useful visual and chemical cues were available *en route*. In addition, inertial cues were presumably obscured by the movement of the boat and constant sloshing of water within the container. However, magnetic cues were not disrupted and might have been perceptible during transport. An additional experiment was therefore conducted to determine whether lobsters could still orient towards the capture site even if all known sensory cues, including magnetic cues, were disrupted during displacement.

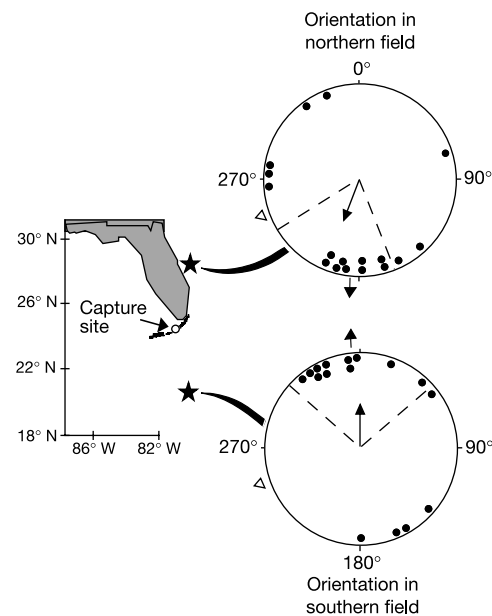


Figure 3 Orientation of lobsters tested in magnetic fields replicating those that exist at two different geographic locations (marked by stars on the map). Lobsters were captured at Grassy Key (CS2 on Fig. 1), transported as before to the testing site (TS1 on Fig. 1), and held overnight in non-magnetic tanks in the local magnetic field. Lobsters tested in a field characteristic of a location north of the test site were significantly oriented ($r = 0.51$, $Z = 4.21$, $P < 0.02$, Rayleigh test) with a mean angle of 199° (upper diagram). Lobsters tested in a field characteristic of a location south of the test site were significantly oriented ($r = 0.43$, $Z = 3.0$, $P < 0.05$, Rayleigh test) but in approximately the opposite direction (a mean angle of 1° ; lower diagram). Other statistical conventions are as in Fig. 1. The arrow outside each orientation diagram indicates the direction in which lobsters would be expected to orient themselves if homing from the fictive locations. The open triangle outside each orientation diagram indicates the actual direction to the capture site from the test site. In each case, lobsters responded as if they had been displaced to the locations marked by the stars rather than by orienting in the direction that was actually towards the capture site.

Lobsters were captured near shore (location CS on Fig. 2), placed in a covered, opaque container and transported overland by lorry to the testing site. In half of the trips the container was lined with an array of magnets, some of which hung from strings and swung erratically, producing fields strong enough to continuously change the alignment of a compass placed anywhere in the container. In the other half of the trips, lobsters were transported in the same container but without magnets. In all trips, the container was suspended by ropes in the cargo section of a lorry so that it swung erratically as the vehicle moved. To disrupt inertial cues further, the lorry was manoeuvred through a series of erratic turns and circles before being driven to the test site 37 km away. All lobsters were then removed from the container and housed overnight in a holding tank in the natural local magnetic field.

In the morning the eyes of the lobsters were covered with rubber caps¹⁷ and their orientation was tested as before. No differences in orientation were found between lobsters that had been transported with magnets and those that had not. Instead, both groups oriented significantly towards the direction of the capture site (Fig. 2). We therefore infer that lobsters deprived of all known orientation cues during transport are nevertheless capable of determining the direction to the capture site when released at unfamiliar locations, and that this homeward direction is apparently determined on the basis of information gathered at the test site. We conclude that lobsters fulfil the criteria of true navigation, as well as the definition of map-based navigation in the new terminology proposed by Able⁷.

How animals capable of true navigation determine their position relative to a goal after displacement into unfamiliar territory is not known. However, recent studies have shown that a few animals can detect elements of the Earth's magnetic field that might, in principle, provide positional information^{18,19}.

Spiny lobsters possess magnetic material that might function in magnetoreception²⁰ and are also known to have a magnetic compass sense¹⁷, but whether they can derive positional information from the Earth's field has not previously been studied. To investigate whether true navigation in lobsters is based on magnetic positional information, lobsters were tested in magnetic fields replicating those that exist at different geographic locations.

Lobsters were collected (location CS2 on Fig. 1) and transported by boat to the first test site (TS1 on Fig. 1) as described for the initial displacement experiments. They were then randomly assigned to two groups: one was tested in a magnetic field that exists at a location ~400 km north of the test site; the other was tested in a field replicating one found at a location ~400 km to the south. Simulated displacements were of greater distance than the actual displacements (Figs 1 and 2) because it is difficult to reproduce consistently the small field differences experienced by lobsters in the initial experiments. However, these simulated displacements are within the range of distances that spiny lobsters are known to move^{21,22}.

Lobsters exposed to the field of the northern site walked south-southwest (Fig. 3). Those exposed to the field of the southern site walked approximately north (Fig. 3). The two distributions are significantly different ($U^2 = 0.269$, $P < 0.01$, Watson test), indicating that lobsters can distinguish between magnetic fields that mark different geographic locations within their environment. Moreover, they responded to each field by orienting in a direction that would have led them towards the capture site had they actually been at the location where each field naturally occurs. These results provide strong evidence that spiny lobsters have a magnetic map sense that is used in navigation.

The precise magnetic feature or features that lobsters detect, and the exact way in which the map is organized, cannot be determined from these initial experiments. The responses to magnetically simulated displacements (Fig. 3) might be explained by proposing that lobsters determine whether they are north or south of a goal by perceiving a single magnetic element (such as inclination or

intensity) that varies in the north-south direction. Such a 'unicoordinate map'¹⁸, however, cannot easily account for the ability to orient towards the capture site from locations that are not along the north-south axis (Figs 1 and 2). Among several alternative possibilities, it might be that lobsters rely on bicoordinate navigation by detecting two different magnetic field elements that vary in different directions across the region², that an unknown non-magnetic cue provides the second coordinate, or that lobsters learn the local magnetic topography and exploit features of the magnetic relief as magnetic 'landmarks'¹¹. Thus, in concluding that lobsters possess a magnetic map, we use the phrase in its most general sense^{9,23}, intending it only as a convenient shorthand to indicate that displaced lobsters can somehow derive sufficient positional information from the Earth's magnetic field to determine the direction towards home. Precisely how they do so remains to be determined.

To use magnetic map navigation over relatively short distances, an animal must overcome several potential problems^{2,10,11}. One requirement is an acute sensitivity to small differences in magnetic fields. A second is an ability to filter out or otherwise compensate for the regular daily variations in the Earth's field, which are caused in part by ionospheric currents^{11,24}. A third is that the magnetic topography in the area must provide a pattern of variation that can be exploited in position-finding.

The degree of magnetic sensitivity in lobsters is not known. However, theoretical considerations suggest that biological receptors can indeed achieve the sensitivity required to use magnetic maps over distances as small as ~10 km (ref. 24). Lobsters are likely to encounter minimal problems with temporal variation in the Earth's field because they normally leave their dens only at night, when the field is most stable¹¹. Analysis of the fine-scale magnetic topography of the study area is not yet possible because detailed magnetic surveys have apparently not been done in this location. However, if magnetic anomalies exist, they do not necessarily impede magnetic navigation because local gradient contours are often aligned in such a way that useful positional information can be extracted from them^{2,11}. Maps of the regional magnetic isolines^{11,24} indicate that, on a larger scale, several magnetic elements vary regularly and predictably across the Florida region and might therefore provide reliable positional information to lobsters or other animals able to detect them.

Regardless of these considerations, our results demonstrate for the first time that an invertebrate animal is capable of true navigation. Moreover, lobsters exposed to magnetic fields replicating those that exist at different geographic locations responded as if homing from each fictive site. These results provide the most direct evidence yet that animals possess and use magnetic maps. Similar mechanisms might function not only in lobsters, but in various animals that migrate or home, including certain fishes²⁵, amphibians¹⁸, reptiles²⁶ and birds²⁷. □

Methods

Animals and study area

Juvenile lobsters (carapace length 47–82 mm) were captured by divers in Florida Bay, USA, a shallow area north and west of the Florida Keys (see map inset and larger map, Fig. 1). Capture site 1 (CS1) was located at Peterson Key (latitude 24.92° N, longitude 80.75° W). Capture site 2 (CS2) was located at Grassy Key (latitude 24.79° N, longitude 80.96° W). Test site 1 (TS1) was located at Long Key (latitude 24.83° N, longitude 80.82° W) and test site 2 (TS2) was located at Pigeon Key (latitude 24.70° N, longitude 81.15° W).

Experimental protocol

The procedure for tethering lobsters has been described previously¹⁷. Each lobster was temporarily blinded with rubber eyecaps, attached by monofilament line to a tracking system modified from one used previously²⁸, and placed into the orientation arena facing in a randomized direction. After a 5-min acclimation period, the data acquisition computer recorded the orientation of the lobster once every 30 s for 30 min. The data were used to calculate a mean angle of orientation for each lobster by using standard procedures for circular statistics²⁹. Those few lobsters that remained motionless for 5 min or longer after being placed into the arena were replaced with other, more active individuals. All trials were conducted during the morning hours after sunrise.

Magnetically simulated displacements

In the experiments described in Fig. 3, the orientation arena was surrounded by a magnetic coil system that was used to control the field in which each lobster walked. The coil system consisted of two different independent four-coil systems arranged orthogonally²⁸. Each coil measured 2.3 m on a side. Lobsters were restricted by a tether to an area in the centre of the coil defined by a horizontal circle of radius 25 cm. In this region, calculated³⁰ and measured deviations from perfect field uniformity were less than 0.5%. Each lobster was then tethered as before and tested in one of two magnetic fields. One field replicated magnetic conditions that exist at a location approximately 400 km to the north, whereas the other replicated a field at a location approximately 400 km to the south. The field used to approximate magnetic conditions at the location north of the test site had an inclination of 59.3° and a total intensity of 47.9 μ T. The field simulating the location south of the test site had an inclination of 51.4° and total intensity of 42.8 μ T. All magnetic field values were verified by three independent measurements with an Applied Physics Fluxgate Magnetometer (model 520A). The experimental fields were based on estimates provided by the International Geomagnetic Reference Field (IGRF) model, 2000 revision, for August 2001 (when the data were collected) using latitude 28.5° N, longitude 80.5° W for the northern site, and latitude 20.5° N, longitude 80.5° W for the southern site. Experiments were conducted in Long Key, Florida (latitude 24.8° N, longitude 80.8° W) where the measured inclination angle was 55.8° and the total field intensity was 45.3 μ T.

Received 6 June; accepted 7 October 2002; doi:10.1038/nature01226.

- Griffin, D. R. Bird navigation. *Biol. Rev.* **27**, 359–400 (1952).
- Phillips, J. B. Magnetic navigation. *J. Theor. Biol.* **180**, 309–319 (1996).
- Phillips, J. B., Adler, K. & Borland, S. C. True navigation by an amphibian. *Anim. Behav.* **50**, 855–858 (1995).
- Wehner, R., Michel, B. & Antonsen, P. Visual navigation in insects: Coupling of egocentric and geocentric information. *J. Exp. Biol.* **199**, 129–140 (1996).
- Walker, M. M. in *Orientation and Communication in Arthropods* (ed. Lehrer, M.) 187–213 (Birkhäuser, Basel, 1997).
- Kramer, G. Experiments in bird orientation and their interpretation. *Ibis* **99**, 196–227 (1957).
- Able, K. P. The concepts and terminology of bird navigation. *J. Av. Biol.* **32**, 174–183 (2001).
- Able, K. P. Common themes and variations in animal orientation systems. *Amer. Zool.* **31**, 157–167 (1991).
- Walcott, C. Pigeon homing: Observations, experiments, confusions. *J. Exp. Biol.* **199**, 21–27 (1996).
- Gould, J. L. The map sense of pigeons. *Nature* **296**, 205–211 (1982).
- Wiltshko, R. & Wiltshko, W. *Magnetic Orientation in Animals* (Springer-Verlag, Berlin, 1995).
- Walcott, C. in *Orientation in Birds* (ed. Berthold, P.) 38–51 (Birkhäuser, Boston, 1991).
- Papi, F. Animal navigation at the end of the century: A retrospect and a look forward. *Ital. J. Zool.* **68**, 171–180 (2001).
- Herrnkind, W. in *The Biology and Management of Lobsters* (eds Cobb, J. S. & Phillips, B. E.) 349–407 (Academic Press, New York, 1980).
- Herrnkind, W., Van Derwalker, J. & Barr, L. Population dynamics, ecology, and behavior of spiny lobster, *Panulirus argus*, of St. John, U. S. Virgin Islands: Habitation and patterns of movements. Results of the Tektite program, vol. 2. *Sci. Bull., Nat. Hist. Mus., Los Angeles Co.* **20**, 31–45 (1975).
- Creaser, E. P. & Travis, D. Evidence of a homing instinct in the Bermuda spiny lobster. *Science* **112**, 169–170 (1950).
- Lohmann, K. J. et al. Magnetic orientation of spiny lobsters in the ocean: Experiments with undersea coil systems. *J. Exp. Biol.* **198**, 2041–2048 (1995).
- Fischer, J. H., Freake, M. J., Borland, S. C. & Phillips, J. B. Evidence for the use of magnetic map information by an amphibian. *Anim. Behav.* **62**, 1–10 (2001).
- Lohmann, K. J. & Lohmann, C. M. F. Detection of magnetic field intensity by sea turtles. *Nature* **380**, 59–61 (1996).
- Lohmann, K. J. Magnetic remanence in the western Atlantic spiny lobster, *Panulirus argus*. *J. Exp. Biol.* **113**, 29–41 (1984).
- Groeneveld, J. C. & Branch, G. M. Long distance migration of South African deep-water rock lobster, *Palinurus gilchristi*. *Mar. Ecol. Prog. Ser.* **232**, 225–238 (2002).
- Gregory, D. R. & Labisky, R. F. Movements of the spiny lobster, *Panulirus argus*, in south Florida. *Can. J. Fish. Aqu. Sci.* **43**, 2228–2234 (1986).
- Gould, J. L. Sensory bases of navigation. *Curr. Biol.* **8**, R731–R737 (1998).
- Kirschvink, J. L., Jones, D. S. & MacFadden, B. J. *Magnetite Biomineralization and Magnetoreception in Organisms* (Plenum, New York, 1985).
- Walker, M. M. et al. Structure and function of the vertebrate magnetic sense. *Nature* **390**, 371–377 (1997).
- Lohmann, K. J., Cain, S. D., Dodge, S. A. & Lohmann, C. M. F. Regional magnetic fields as navigational markers for sea turtles. *Science (Washington DC)* **294**, 364–366 (2001).
- Munro, U., Munro, J. A., Phillips, J. B., Wiltshko, R. & Wiltshko, W. Evidence for a magnetite-based navigational “map” in birds. *Naturwissenschaften* **84**, 26–28 (1997).
- Lohmann, K. J. & Lohmann, C. M. F. Detection of magnetic inclination angle by sea turtles: A possible mechanism for determining latitude. *J. Exp. Biol.* **194**, 23–32 (1994).
- Zar, J. H. *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, New Jersey, USA, 1996).
- Kirschvink, J. Uniform magnetic fields and double wrapped coil systems: Improved techniques for the design of bioelectromagnetic experiments. *Bioelectromag.* **13**, 401–411 (1992).

Acknowledgements We thank K. Cohen, M. Piet and S. Peters for research assistance, and C. Lohmann and S. Johnsen for a critical reading of manuscript drafts. This work was supported by grants from the PADI Foundation and the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.C.B. (e-mail: LBoles@email.unc.edu).

Role of duplicate genes in genetic robustness against null mutations

Zhenglong Gu^{*†}, Lars M. Steinmetz^{‡†}, Xun Gu[§], Curt Scharfe[‡], Ronald W. Davis[‡] & Wen-Hsiung Li^{*}

^{*} Department of Ecology & Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA

[‡] Department of Biochemistry, and Stanford Genome Technology Center, Stanford University School of Medicine, Stanford, California 94305, USA

[§] Departments of Zoology and Genetics, Iowa State University, Ames, Iowa 50011, USA

[†] These authors contributed equally to this work

Deleting a gene in an organism often has little phenotypic effect^{1–5}, owing to two mechanisms of compensation^{4–10}. The first is the existence of duplicate genes: that is, the loss of function in one copy can be compensated by the other copy or copies. The second mechanism of compensation stems from alternative metabolic pathways, regulatory networks, and so on. The relative importance of the two mechanisms has not been investigated except for a limited study, which suggested that the role of duplicate genes in compensation is negligible¹⁰. The availability of fitness data for a nearly complete set of single-gene-deletion mutants of the *Saccharomyces cerevisiae* genome¹¹ has enabled us to carry out a genome-wide evaluation of the role of duplicate genes in genetic robustness against null mutations. Here we show that there is a significantly higher probability of functional compensation for a duplicate gene than for a singleton, a high correlation between the frequency of compensation and the sequence similarity of two duplicates, and a higher probability of a severe fitness effect when the duplicate copy that is more highly expressed is deleted. We estimate that in *S. cerevisiae* at least a quarter of those gene deletions that have no phenotype are compensated by duplicate genes.

No correlation was found between the sequence similarity of duplicate genes and the fitness effect of a null mutation in one of the two duplicates when functional data from the yeast *S. cerevisiae* was analysed previously¹⁰. It was therefore concluded that gene duplications contribute little to the ability of an organism to withstand mutations (genetic robustness), although they may be responsible for a small fraction of weak, null-mutation phenotypes¹². Because this conclusion was based on only 45 duplicate genes, however, the issue deserves further investigation. Indeed, this conclusion is not supported by a limited analysis of a third of the genes in the yeast genome¹ and is contrary to the general observation of relaxed selective constraints after gene duplication^{13,14}.

From 5,766 yeast open reading frames (ORFs) for which we had a fitness measure of strains with a corresponding single-gene deletion¹¹, we found 1,509 duplicate (paralogous) genes. To avoid including pseudogenes and erroneously predicted genes, we subsequently analysed only genes that had been studied previously (that is, each had a gene name in the *Saccharomyces* Genome Database (SGD) in addition to its ORF name). This yielded 1,275 singleton genes, and 1,147 duplicate genes that had at least one paralogue elsewhere in the genome. We compared the frequency distribution of fitness for duplicate genes with that for singletons (Fig. 1a). We classified genes into four groups on the basis of the minimum fitness value for a strain across the five different growth conditions tested (Methods) including both fermentation and respiration, the main growth conditions of yeast.

The two distributions were significantly different ($P \ll 0.001$): duplicate genes had a significantly lower proportion of genes with a lethal effect of deletion (12.4% versus 29.0%) and a significantly higher proportion of genes with a weak or no effect of gene deletion

(64.3% versus 39.5%, the latter value is similar to a previous estimate of 43.6% (ref. 10); Fig. 1a). This comparison indicates that there is a significantly higher probability of functional compensation for a duplicate gene than for a singleton. We emphasize that 'compensation' here does not imply that the gene is dispensable in long-term evolution but means that the gene is dispensable in an individual under the conditions tested⁶. When the genes with a lethal effect of deletion were excluded from both collections, the difference between the two distributions remained significant ($P \ll 0.001$; Supplementary Table 1).

To see whether the above conclusion held for different growth environments, we compared the distributions of fitness (f) for duplicate and singleton genes under each of the five growth conditions studied¹¹. The empirical cumulative distributions of f under the YPD growth condition (Methods) for duplicate genes and for singletons are shown in Fig. 1b. The Kolmogorov–Smirnov test indicated that the two distributions were significantly different ($P = 2.2 \times 10^{-16}$). The same conclusion held for the other four growth conditions (Supplementary Fig. 1).

If duplicate genes tend to compensate for each other's function, then a testable prediction is that deletions of duplicate genes should tend to show similar fitness effects. To avoid multiple comparisons within a gene family, independent (non-overlapping) duplicate gene pairs were selected randomly from each gene family. For each of the 418 duplicate gene pairs selected (with both copies having been previously studied), we computed the difference between the fitness effects of duplicate genes i and j (D_{ij}) and then obtained the mean of all D_{ij} values (D^*) for each growth condition. For comparison, 418 protein pairs were selected randomly from all previously studied genes and the D^* value was calculated as above. This procedure was repeated 100,000 times to

derive a frequency distribution of the mean difference in fitness effects between genes for each studied condition. The mean value ($D^* = 0.193$) for duplicate genes was far lower than the mean value for random gene pairs ($P < 10^{-5}$) under the YPD growth condition (Fig. 2), confirming that duplicates tend to show more similar fitness effects of deletion than do random gene pairs. The same results held for the four other growth conditions (Supplementary Fig. 2).

The two genes derived from a duplication should initially have the same function. In long-term evolution, the accumulation of mutations in both copies over time results in either functional loss in one copy or functional divergence between the two copies¹⁵. If gene duplication is important for genetic robustness, genes with close paralogues should be compensated for deletion more often than genes with only distant paralogues. To test this hypothesis, we focused on a set of duplicate genes that excluded ribosomal proteins because of their unusual properties, such as strong codon usage bias, very high expression, and severe fitness effects of null mutations. This set of duplicate genes was divided further into different groups on the basis of the nonsynonymous distance (K_A) of each gene to its most similar gene in the genome (defined as the gene with the smallest K_A value to the studied duplicate gene). Within each K_A interval, the frequencies of genes with different values of f were calculated. The fitness classification was based on the minimum fitness of a strain across all of the five tested media conditions, as in Fig. 1a. As expected under the above hypothesis, the proportion of genes with a weak or no effect decreased with K_A (correlation coefficient, $R = -0.95$; $P < 0.001$), whereas the proportion of genes that are lethal when deleted increased with K_A ($R = 0.94$, $P < 0.001$; Fig. 3). This observation is contrary to the previous conclusion that there is no correlation between K_A and the fitness effect of deleting a duplicate gene, which was based on a much smaller data set¹⁰.

As the sequence similarity between duplicated genes decreases, their frequency of compensation will approach that for singletons. But even among duplicate genes with a K_A greater than 0.7 from their most similar paralogues in the whole genome, about 53% still had a weak effect or no effect when deleted (Fig. 3), which was significantly higher than the 39.5% of singletons that showed a weak or no effect of deletion (Fig. 1a, χ^2 -test, $P < 0.01$), implying that the compensation effect might exist even for ancient duplicate genes. Nevertheless, the decreasing compensation effect between duplicate genes with K_A suggests that the functional divergence of duplicate genes is coupled to some extent with the divergence of their protein sequences.

Because expression level is used frequently to infer the activity and function of gene products, we tested whether deleting the copy

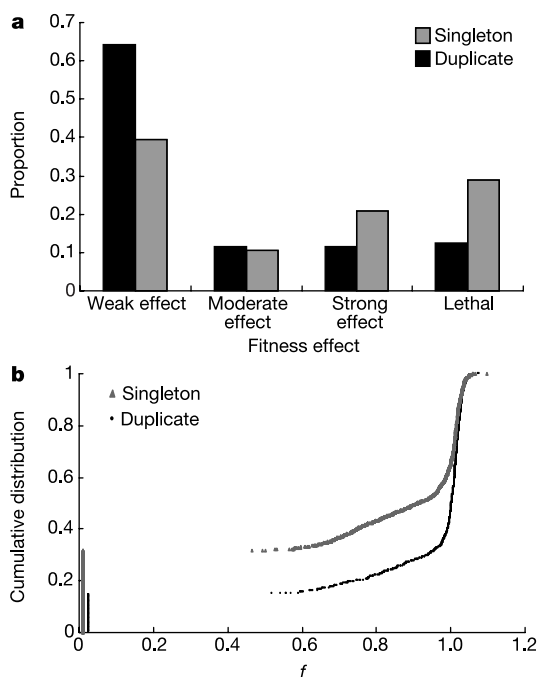


Figure 1 Distributions of fitness (f). **a**, Discrete distributions of f for singleton genes and for duplicate genes. The difference for the two distributions is significant (a contingency table test, $P \ll 0.001$) under YPD growth conditions. **b**, Empirical cumulative distributions of f for singleton genes and for duplicate genes. Data points with $f = 0$ have been shifted to $f = 0.025$ for duplicate genes and to $f = 0.01$ for singleton genes to distinguish between the two curves, but remained unchanged in the statistical analysis. The Kolmogorov–Smirnov test shows that the two distributions are significantly different ($P = 2.2 \times 10^{-16}$).

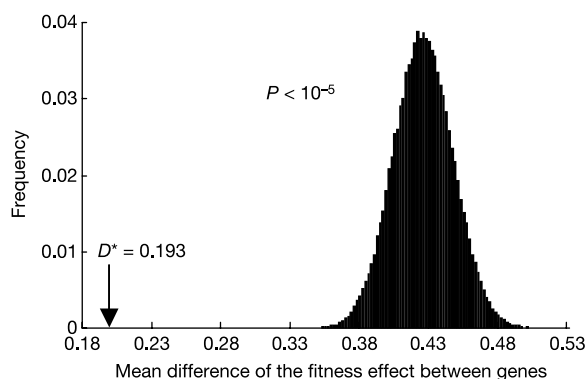


Figure 2 Distribution of mean fitness differences between randomly selected gene pairs (100,000 replicates each with 418 gene pairs) under the YPD growth condition. Arrow indicates the mean difference ($D^* = 0.193$) between duplicate genes.

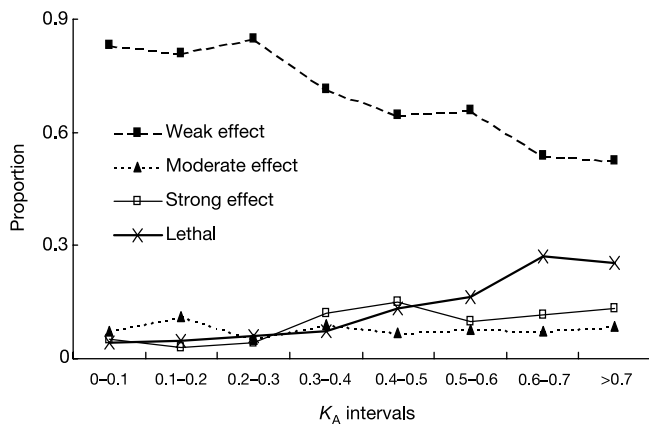


Figure 3 Relationship between protein distance and fitness effect of deletion. Protein distance is measured by the K_A of each gene to its most similar paralogue in the genome. The proportion of genes with a weak effect of deletion decreases with K_A ($R = -0.95$, $P < 0.001$), whereas the proportion of genes with lethal effect increases with K_A ($R = 0.94$, $P < 0.001$).

of a duplicate gene that is more highly expressed would have a stronger fitness effect than deleting the copy with lower expression. We considered only duplicate gene pairs with different fitness effects of deletion (that is, those in which the relative fitness difference (as defined by $|(f_i - f_j)/(f_i + f_j)/2|$) for genes i and j was larger than 5%, or those for which one of the two duplicate copies is essential and the other is not). Expression (absolute transcript abundance) was estimated using available data measured by Affymetrix microarray experiments¹⁶. We used the fitness effects of the duplicate genes measured under the YPD growth condition in this analysis because the expression levels were also measured under the YPD growth condition¹⁶. Deleting the duplicate gene that has higher expression indeed tended to have a larger fitness effect (Table 1). For example, in 72 of the 98 gene pairs where the deletions had different fitness effects, the stronger fitness effect was seen in the more highly expressed gene.

The high frequency of genes that have weak or no fitness effects of deletion among single genes, as well as among duplicate genes (Fig. 1a), indicates that the fitness effect of a gene deletion is also affected by factors other than copy redundancy. Whereas some genes may show null-mutation phenotypes only under experimental conditions that differ from the five growth conditions tested here, a fraction of weak, null-mutation phenotypes among singletons might be due to compensation through alternative pathways or network branching, as suggested previously¹⁰. The relative importance of the compensating mechanism through functionally redundant duplicate genes can be estimated roughly as follows. If we assume that the extra proportion of genes with a weak or no fitness effect of deletion in duplicate genes when compared with the proportion for singleton genes is due to copy redundancy (64.3% for duplicates, 39.5% for singletons; difference 24.8%; Fig. 1a), this will give the lower bound of the contribution of gene duplication to genetic robustness. The proportion of this contribution (G) is estimated to be 23% because for 284 genes (1,147 duplicates $\times$ 24.8%) out of a total of 1,241 genes that are robust against deletion (1,147 duplicates $\times$ 64.3% + 1,275 singletons $\times$ 39.5%), the robustness can be attributed to gene duplication. The upper bound can be estimated by assuming that all of the genes with a weak or no fitness effect of deletion in duplicate genes are due to copy redundancy. This gives an upper estimate of 59% for G because 738 duplicate genes (1,147 duplicates $\times$ 64.3%) and 503 singleton genes (1,275 singletons $\times$ 39.5%) showed a weak or no fitness effect of deletion (Fig. 1a).

As our estimates of G were based on only about a third of all genes

Table 1 Relationship between expression and fitness effect of null mutation in duplicate genes*

Relative expression	Number of genes with a larger fitness effect of deletion		
	Different effect†	One lethal	Similar effect‡
High	72‡	50‡	125
Low	26	12	108
Total	98	62	233

*Only gene pairs for which both copies have been studied previously were included.

†Two duplicate genes i and j were said to have different fitness effects if $|(f_i - f_j)/(f_i + f_j)/2| > 0.05$, but similar fitness effects if otherwise.

‡Significant at $P < 0.001$.

in the yeast genome owing to the stringent criteria used originally to define singletons and duplicate genes (Methods), we repeated the calculations on more genes by using a less stringent criterion to group genes. We divided the whole set of yeast genes into singletons and duplicate genes as follows: a protein was defined as a singleton if it did not have a hit with any other proteins in a FASTA¹⁷ search of the whole yeast proteome (6,357 ORFs) with $E = 0.01$ (E is the expected number of hits by chance); otherwise, it was a duplicate gene. This procedure led to 3,249 duplicate genes and 3,108 singletons. We then selected only genes that have been studied previously: 2,240 duplicate genes and 1,567 singletons, for which 2,063 and 1,477, respectively, had an associated fitness measurement. By applying the above procedure to these two sets of duplicate genes and singletons, we estimated G to be between 0.21 and 0.67; these two values are not very different from the above lower and upper bounds, respectively. In conclusion, the analyses in this study provide strong evidence for the importance of duplicate genes in genetic robustness against null mutations.

Although our estimates are compatible with the view that interactions among unrelated genes rather than duplicate genes are the main cause of genetic robustness against mutations^{10,18}, two additional factors need to be considered. First, because we have considered only five growth conditions, it is possible that when a gene deletion showed no effect in any of these conditions it was not due to compensation by other genes but was because the gene deleted was not related to the growth conditions used. Intuitively, when more growth conditions are studied, both the proportion of duplicate genes and the proportion of singletons that show only a weak or no effect of deletion on growth rate will decrease. Indeed, the two proportions were 70.9% and 49.2% when only the YPD growth condition was considered (data not shown), but became 64.3% and 39.5% when the five growth conditions shown in Fig. 1a were used. The decrease is larger for singletons than for duplicate genes, probably because duplicate genes have on average a stronger overlap in function than do singletons and so can compensate each other in a wider range of conditions. For this reason, our lower bound of 23% for the relative contribution of duplicate genes to compensation for null mutations is likely to be an underestimate. Second, a singleton in this or other studies could actually have one or more paralogues in the genome that cannot be detected by the criteria used but still overlap in function. Thus, gene duplication might be the ultimate origin of functional compensation for some 'singletons'. In conclusion, whether the contribution of gene duplication to genetic robustness is really less important than interactions among unrelated genes is an issue that remains to be resolved by further studies. □

Methods

Fitness measurements

Fitness measurements were obtained from a high-throughput study¹¹ that measured the growth of each strain of a nearly complete collection of yeast single-gene-deletion mutants under both fermentable and non-fermentable (respiratory) growth conditions. We studied five growth media: YPD (1% Bacto-peptone (Difco), 2% yeast extract and 2% glucose), YPDGE (0.1% glucose, 3% glycerol and 2% ethanol), YPE (2% ethanol), YPG (3% glycerol) and YPL (2% lactate). Each strain contained the precise homozygous

diploid deletion of 1 of 4,706 ORFs in the yeast genome. We calculated the fitness values for each media condition as the extent of survival and reproduction (fitness) of the deletion strain relative to the pool of all strains grown and measured collectively. Fitness values (f) of 1.0 indicate no difference between a single strain and the pool average for that condition; $f < 1.0$ indicates that the strain is less fit, and $f > 1.0$ indicates that the strain is more fit than the pool average. In addition, we added to our analysis 1,060 ORFs that each had a lethal effect when deleted and assayed in YPD; we used only lethal deletions that could be inferred as lethal from both of the two studies conducted^{11,19}. We divided all genes into four groups according to the f value as follows: (1) if $f > 0.95$ for all five media conditions¹¹, the deletion has a weak or no fitness effect in all conditions; (2) if $0.8 \leq f_{\min} < 0.95$, where $f_{\min}$ is the smallest f value for all five growth conditions, the deletion has a moderate effect; (3) if $0 < f_{\min} < 0.8$, the deletion has a strong effect; and (4) if the deletion is lethal, we set $f = 0$.

Identification of duplicate and singleton genes

An all-against-all FASTA¹⁷ search was conducted for the whole set of *S. cerevisiae* protein sequences (downloaded from SGD, <http://genome-www.stanford.edu/Saccharomyces/>). A single-copy gene (that is, a singleton) was defined as a protein that did not hit any other proteins in the FASTA search with $E = 0.1$; this loose similarity search criterion was used to make sure that a singleton is indeed a singleton. (An even looser criterion $E = 1$ was also used in the definition of singleton genes but the results were essentially the same.) Duplicate genes were identified as described²⁰ except that the criterion of 80% alignable regions between protein sequences was reduced to 50%, because the 80% requirement can miss some duplicate genes. This relaxed criterion is still conservative for identifying duplicate genes; but because we wanted to detect the differences in fitness effects between real duplicate genes and singletons, we used stringent criteria to define duplicate genes and single genes. For each protein pair that met the homology criteria, the FASTA alignable regions were realigned using ClustalW²¹ and the corresponding coding sequences were aligned on the basis of the protein alignments. The number of substitutions per synonymous site (K_S) and the number of substitutions per nonsynonymous site (K_A) between duplicate genes were estimated by the PAML package²² using default parameters.

Estimation of gene expression

The Affymetrix microarray gene expression data were downloaded from a study investigating gene expression in response to environmental changes across the genome¹⁶. As controls, gene expression was measured at time zero in YPD before each environmental change was conducted. We averaged seven independent time-zero measurements to obtain an estimate of the absolute abundance of each messenger RNA transcript in YPD. Our data showed a very high correlation ($r = 0.95$, $P = 0$) to previous estimates of the absolute abundance of all transcripts in the yeast genome during YPD growth²³.

Received 29 July; accepted 16 September 2002; doi:10.1038/nature01198.

- Winzler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
- Cadigan, K. M., Grossniklaus, U. & Gehring, W. J. Functional redundancy: the respective roles of the 2 sloppy paired genes in *Drosophila* segmentation. *Proc. Natl Acad. Sci. USA* **91**, 6324–6328 (1994).
- Saga, Y., Yagi, T., Ikawa, Y., Sakakura, T. & Aizawa, S. Mice develop normally without tenascin. *Genes Dev.* **6**, 1821–1831 (1992).
- Gibson, T. J. & Spring, J. Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. *Trends Genet.* **14**, 46–49 (1998).
- Normanly, J. & Bartel, B. Redundancy as a way of life—IAA metabolism. *Curr. Opin. Plant Biol.* **2**, 207–213 (1999).
- Brookfield, J. F. Y. Can genes be truly redundant? *Curr. Biol.* **2**, 553–554 (1992).
- Nowak, M. A., Boerlijst, M. C., Cooke, J. & Smith, J. M. Evolution of genetic redundancy. *Nature* **388**, 167–171 (1997).
- Tautz, D. Redundancies, development and the flow of information. *BioEssays* **14**, 263–266 (1992).
- Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
- Wagner, A. Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24**, 355–361 (2000).
- Steinmetz, L. M. *et al.* Systematic screen for human disease genes in yeast. *Nature Genet.* **31**, 400–404 (2002).
- Seoighe, C. & Wolfe, K. H. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* **2**, 548–554 (1999).
- Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
- Conant, G. C. & Wagner, A. GenomeHistory: a software tool and its application to fully sequenced genomes. *Nucleic Acids Res.* **30**, 3378–3386 (2002).
- Ohno, S. *Evolution by Gene Duplication* (Springer, Berlin, 1970).
- Causton, H. C. *et al.* Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **12**, 323–337 (2001).
- Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence analysis. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
- Kitami, T. & Nadeau, J. H. Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nature Genet.* **32**, 191–194 (2002).
- Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
- Gu, Z., Cavalcanti, A., Chen, F.-C., Bouman, P. & Li, W.-H. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol. Biol. Evol.* **19**, 256–262 (2002).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

- Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
- Holstege, F. C. P. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank M. Long, T. Oakley, D. Nicolae, K. Thornton, A. Deutschbauer and J. Zhang for discussions and comments. This work was supported by NIH grants to W.H.L. and to R.W.D.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.H.L. (e-mail: whli@uchicago.edu).

Synaptic depression in the localization of sound

Daniel L. Cook*, Peter C. Schwindt*, Lucinda A. Grande† & William J. Spain*†‡§

* Department of Physiology and Biophysics, † Department of Neurology; and § Virginia Merrill Bloedel Hearing Research Center, University of Washington, Seattle, Washington 98105, USA

† VA Puget Sound Health Care System, Seattle, Washington 98108, USA

Short-term synaptic plasticity, which is common in the central nervous system, may contribute to the signal processing functions of both temporal integration and coincidence detection^{1–3}. For temporal integrators, whose output firing rate depends on a running average of recent synaptic inputs, plasticity modulates input synaptic strength and thus may directly control signalling gain² and the function of neural networks^{1–4}. But the firing probability of an ideal coincidence detector would depend on the temporal coincidence of events rather than on the average frequency of synaptic events. Here we have examined a specific case of how synaptic plasticity can affect temporal coincidence detection, by experimentally characterizing synaptic depression at the synapse between neurons in the nucleus magnocellularis and coincidence detection neurons in the nucleus laminaris in the chick auditory brainstem⁵. We combine an empirical description of this depression with a biophysical model of signalling in the nucleus laminaris. The resulting model predicts that synaptic depression provides an adaptive mechanism for preserving interaural time-delay information (a proxy for the location of sound in space) despite the confounding effects of sound-intensity-related information. This mechanism may help nucleus laminaris neurons to pass specific sound localization information to higher processing centres.

In the avian auditory brainstem, each neuron of the nucleus laminaris (NL), like the medial superior olive in mammals^{6,7}, constitutes an azimuthal 'place code'⁸ that is tuned to detect the coincidental arrival of excitatory postsynaptic potentials (EPSPs) generated by ipsi- and contralateral nucleus magnocellularis (NM) neuron action potential firing⁹, which is phase-locked to sound waves arriving at each ear. The intensity of the sound, however, changes the likelihood that an NM cell will fire on a given sound wave cycle, such that increasing sound intensity increases the average NM firing rate and thus the average frequency of NM–NL EPSPs¹⁰. This increase of NM cell firing rate with sound intensity would be expected to increase the postsynaptic NL cell firing rate,

perhaps even to the saturation limit of its spike-generating mechanism. Because NL cell firing rate encodes sound localization⁸, the saturation of its firing rate will degrade sound localization.

The NL cell firing rate in the owl is remarkably insensitive to sound intensity¹¹, however, suggesting that there are mechanisms that minimize the effect of intensity-related inputs to higher centres processing the localization of sound. Three such mechanisms have been postulated: first, the intrinsic membrane properties of chick NL coincidence detectors may limit their firing rate¹²; second, parallel, intensity-related inhibitory afferents to chick and owl NL cells may attenuate the NL responses to high-intensity sound^{11,13–15}; and third, nonlinear, saturating summation of EPSPs in NL neuron dendrites may limit excessive input¹⁶. Here we suggest that the observed synaptic depression at NM–NL synapses may be an additional, independent mechanism for enhancing coincidence detection (and thus sound localization) in the auditory system.

We studied synaptic depression at NM–NL synapses in brainstem slices taken from chicks at embryonic day E18–21 (hearing in chick embryos is established at about E15–17; ref. 17). Figure 1a shows how we used minimally suprathreshold extracellular stimulation of identified NM neurons to elicit EPSPs that were recorded intracellularly from ipsilateral NL neurons. Figure 1b shows an example recording of a 50-Hz stimulus train to an NM neuron in which the EPSP amplitude depressed quickly to a steady-state value much

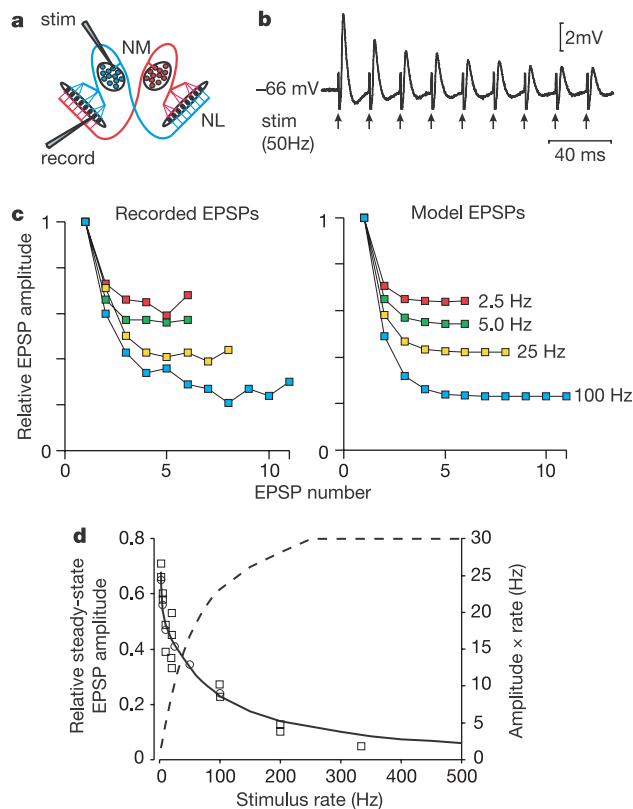


Figure 1 Synaptic depression at NM–NL synapses. **a**, EPSPs were recorded from NL neuron somata in chick brainstem slices in response to minimal stimulation (stim) of ipsilateral NM cells (duration 0.2–0.6 ms, amplitude 20–50 μ A). **b**, EPSP amplitudes rapidly decrease on successive stimuli. **c**, Plots of average EPSP amplitude versus EPSP number (15–20 stimulus trains at 2–100 Hz with intertrain intervals of 5–10 s) Left, measured; right, simulated from equation (4). Each EPSP amplitude was normalized to the first EPSP in its train. **d**, Steady-state EPSP amplitude decreases with increasing stimulus rate in a manner closely predicted by the empirical model of synaptic depression (unbroken line). Circles represent data from the cell in **c**; squares represent data from four other cells. The broken line shows that the product of EPSP amplitude times stimulus rate (right axis) reaches a maximum at about 250 Hz and thus defines a ‘limiting frequency’^{2,21}.

smaller than the first EPSP. Figure 1c (left) shows the time course of depression at each of four stimulus frequencies for one set of NL cell recordings. Successive EPSPs declined to a steady-state amplitude after five or six EPSPs. The steady-state amplitude was reduced to roughly a third as the stimulus frequency increased from 2 to 100 Hz.

To test the idea that the observed synaptic depression at NM–NL synapses may preserve coincidence detection in NL neurons during increased NM neuron firing, we used NEURON software¹⁸ to model the neuronal circuit shown in Fig. 1a. The model consisted of three key elements: first, synaptic inputs from several ipsi- and contralateral NM neurons converged on a single NL cell; second, an empirically derived model simulated NM–NL synaptic depression (Methods and Fig. 1c, d); and third, a biophysical, compartmental model simulated the unusual EPSP summation and spiking properties of NL cells¹².

To provide realistic synaptic inputs to NL neurons, we used a stimulus model^{10,11,19} that mimics NM neuron firing patterns (Fig. 2). When the sound stimulus is a pure tone (a sinusoidal variation in air pressure), action potentials in a given NM neuron tend to fire at a specific time delay (or phase delay) after the peak of each cycle of the tone stimulus. Such ‘phase-locking’ to the audio signal is not precise, however, because there is a small phase error, called ‘jitter’, in spike timing, which is evident as scatter around the mean in the histograms shown in Fig. 2c. In addition, NM cells fire on only a fraction of sound cycles so that their average firing frequency is always less than the tone frequency (Fig. 2b). But increased sound intensity increases the probability of NM firing, so that the average firing frequency amounts to a proxy variable for sound intensity¹⁰.

In the computational model, the phase-locked, jittery and prob-

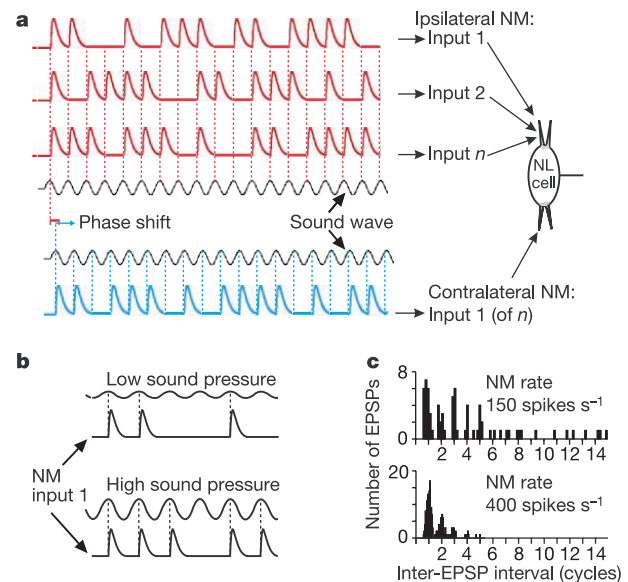


Figure 2 Simulating phase-locked NM synaptic input to NL neurons. **a**, A pure tone (upper sound waves) arriving at the ipsilateral ear elicits phase-locked NM–NL EPSPs (red traces represent 3 of 72 NM inputs) that converge on one set of dendrites on an NL neuron. The same tone arriving with some phase delay (lower sound wave) at the contralateral ear triggers phase-delayed trains (blue trace represents 1 of 72 inputs) of EPSPs that converge on a second set of dendrites on the NL neuron. The number of EPSPs arriving at the NL cell varies from cycle to cycle, because each NM cell fires once on only a fraction of the sound cycles according to a probability distribution. Included in the model but not shown is the ‘jitter’ of NM firing. **b**, Increasing sound intensity increased the probability per cycle of NM cell firing^{10,11}. **c**, An increase in average NM firing rate from 150 to 400 s^{-1} (that is, a change from low to high sound intensity) causes a shortening of the average inter-EPSP interval.

abilistic firing of two sets of 72 independent synapses (one set from ipsi- and one set from contralateral NM cells) converge on opposite sets of NL dendrites. On successive sound cycles, the NL neuron receives a random number of EPSPs because of the probabilistic nature of the NM cell firing. We introduced a phase offset (from 0 to 180°; Fig. 2a) between ipsi- and contralateral NM cell spike trains to represent the interaural time difference (ITD; up to 83 μ s for a 600-Hz virtual sound) occasioned by an off-centre sound source. At a phase offset of 0° (corresponding to an on-centre sound source), ipsi- and contralateral inputs from the NM neurons arrive coincidentally on the NL neuron and thus maximize the likelihood of depolarizing the NL neuron to its spike threshold to evoke its maximal firing rate. Increasing the phase delay towards 180° decreases the likelihood of ipsi- and contralateral EPSPs from NM neurons arriving coincidentally, so that the average NL firing rate decreases to a minimum^{8,11}.

A synaptic depletion model^{20,21} fit well to the observed NM–NL synaptic depression (Fig. 1c, right, and Fig. 1d, unbroken line). Because we did not study the mechanisms underlying NM–NL synaptic depression, we do not claim that depletion accounts for depression at the NM–NL synapse, but use this model simply as a convenient and accurate empirical description of the time course of the observed synaptic depression.

The time course of the postsynaptic conductance increase in the NL dendrites caused by each NM synapse was modelled as an alpha function (see equations (1) and (2) in Methods) with a time constant of 0.26 ms (ref. 12). In runs without synaptic depression, the maximal conductance amplitude ($G_{\max}$) was fixed and independent of NM firing rate. In runs using synaptic depression, $G_{\max}$ was applied only to the first NM action potential after the onset of a tone, after which synaptic conductance on successive EPSPs was computed by the depletion model (refs 20, 21, and see below).

To simulate the NL cell response to NM–NL EPSPs, we modelled the biophysics of NL cells (Methods) according to our previous findings¹² that, first, unlike classical integrating neurons, NL cells fire only a single initial action potential in response to sustained depolarization (Fig. 3, bottom); and second, NL cells fire reliably during suprathreshold low-frequency input events (Fig. 3, top) but miss events at higher frequencies (Fig. 3, middle).

Our model coincidence detector circuit showed the expected property of a smooth reduction of NL firing rate as phase delay increased from 0 to 180° (Fig. 4a, left). This behaviour was apparent over a range of synaptic conductance values and in the absence of synaptic depression, as long as the average NM cell firing rate (the proxy for sound intensity) was low. But we found that no single value of $G_{\max}$ could provide coincidence detection at both high and low sound intensities. For example, a $G_{\max}$ of 1.5 or 2.0 nS worked well at low intensities but offered virtually no discrimination of phase delay (sound location) for the first 90° at high sound intensity (Fig. 4a, right, upper curves). Conversely, a $G_{\max}$ of 0.5 nS provided coincidence detection at high intensity but no discrimination at low intensity (Fig. 4a, lowest curves).

The above results indicate that $G_{\max}$ needs to be varied to provide coincidence detection at both low and high NM firing rates (sound intensity). Synaptic depression can provide an automatic mechanism for adjusting $G_{\max}$ inversely to the NM firing rate to provide roughly the same NL neuron coincidence detection irrespective of sound intensity. This fact is illustrated by Fig. 4b, which shows that the NL phase-delay curves with synaptic depression are tightly clustered for a broad range of sound intensities represented by average NM firing rates from 150 to 450 spikes s^{-1} . With synaptic depression included in the model, the NL firing rate becomes relatively insensitive to changes in sound intensity, as is observed *in vivo*¹¹. Thus, there is a comparable degree of coincidence detection at all sound intensity levels by automatically adjusting the average synaptic conductance inversely to the average input frequency. When the average firing rate of the inputs is greater than the

limiting frequency, the NL neuron behaves like an ideal coincidence detector (that is, sound intensity has almost no effect on firing rate).

Developmental studies have revealed that some synapses show less depression with maturation (for example, see ref. 22). We therefore recalculated the NL firing rate versus phase-delay curves with $U = 0.2$ and $G_{\max} = 3.3$ nS (see equation (4) in Methods) to reduce the amount of steady-state depressed synaptic conductance to about 50% of the value that we measured experimentally. The results were similar to those shown in Fig. 4b, in that we obtained non-saturating phase tuning curves over a wide range of sound intensities (data not shown).

The underlying depression of synaptic conductance going from low- to high-intensity sounds is evident in Fig. 4c, where the average synaptic conductance decreased from 1.4 to 0.7 nS as the input average NM firing rate increased from 150 to 400 spikes s^{-1} . The synaptic conductance decrease during depression, and not some other consequence of the depression kinetics, accounts for the changes in NL cell firing rate, because the coincidence detection curves with depression for 150 and 400 spikes s^{-1} can be overlaid almost exactly by the curves without synaptic depression in which $G_{\max}$ was adjusted to 1.4 nS and 0.7 nS (Fig. 4d). These values correspond to the average steady-state conductances attained during depression (Fig. 4c).

Our data show that synaptic depression exists at the NM–NL synapse in auditory signal processing neurons of the avian brainstem. The fast component of the depression kinetics (tens of milliseconds) is significantly faster than those found in most other areas of the brain (hundreds of milliseconds)^{2,21}, which perhaps reflects the rapidly firing cells present in the hippocampus^{23,24} and the auditory signal processing system^{22,25}.

We neither sought nor can we offer a mechanism for the depression that we have described. Instead, we were interested in understanding what role, if any, synaptic depression may have in signal processing according to the model of auditory coincidence detection and sound localization²⁶. Towards this goal, we built a computational model to examine whether the measured features of synaptic depression can offer improved sound localization. Our premise is that sound localization by higher centres (beyond NM and NL cells) would benefit from nearly pure coincidence detection by the NL cell, as would be the case if NL cell firing depended primarily on the degree of coincidence of the synaptic events and

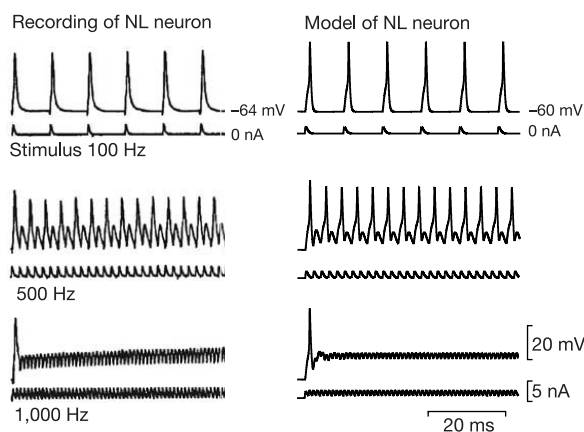


Figure 3 Distinctive firing properties of chick and model NL neurons. Chick neuron is shown on the left (reproduced, with permission, from ref. 12); model neuron is shown on the right. Top, repetitive suprathreshold stimulation at a low frequency (100 Hz) consistently evokes action potentials. Middle, high-frequency (500 Hz) stimulation fails to trigger action potentials on alternate cycles. Bottom, very high-frequency (1,000 Hz) or direct current (DC) stimulation evokes only a single initial action potential. The ability of these neurons to filter out temporally summated inputs is well suited to their function as coincidence detectors¹², in contrast to integrating neurons, which fire repetitively to DC inputs.

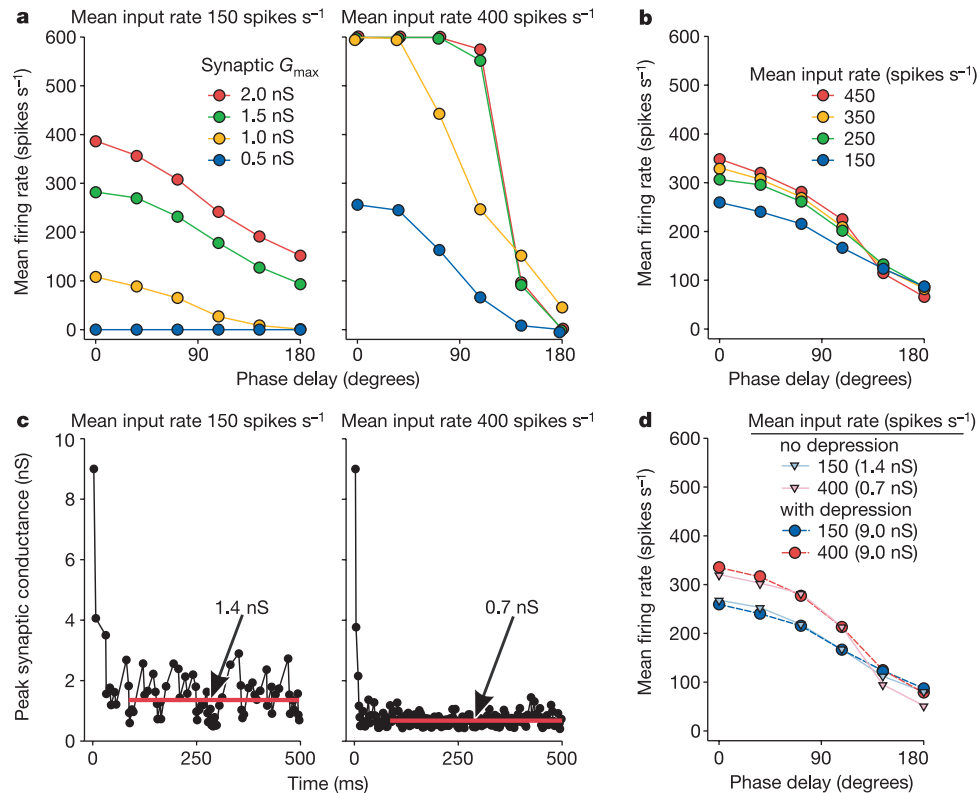


Figure 4 Synaptic depression buffers coincidence detection against changing sound intensity. **a**, Without synaptic depression, at low sound intensity (average NM input firing rate 150 s⁻¹, left) with $G_{max} < 1.0$ nS, NL output firing depends only weakly on phase delay; but at high sound intensity (400 s⁻¹, right) with $G_{max} > 1.0$ nS, firing rates saturate for phase delays <90° and decline to near zero for phase delays near 180°. Curves for low and high intensity are, however, nearly identical for $G_{max} = 1.5$ and $G_{max} = 0.5$ nS, respectively. **b**, With synaptic depression, coincidence detection curves converge and, above the limiting frequency (250 s⁻¹, Fig. 1d), have little dependence on

input firing rate. Note that synaptic depression also minimizes the reduction in firing rate at long phase delays as sound intensity increases. **c**, Variation of peak event conductance in a single depressing NM-to-NL synapse during a low-intensity (left) and a high-intensity (right) tonal stimulus. Depression reduced average peak synaptic conductance from 1.4 to 0.7 nS as input firing rate increased from 150 to 400 s⁻¹. **d**, Preventing event-to-event depression kinetics (apparent in **c**) by fixing G_{max} at 1.4 and 0.7 nS to correspond to firing rates of 150 and 400 s⁻¹, respectively, has no effect on coincidence detection. The sound frequency is 600 Hz in **a–d**.

less so on the frequency of coincidences. Our model shows that synaptic depression can indeed provide an adaptive mechanism that may contribute to the remarkable resistance of NL phase-related spiking (sound localization) to changes of NM cell input spike frequency (sound intensity).

Experimental and theoretical studies have emphasized that synaptic depression allows cortical neurons to function as signal differentiators where the transient response to changes of input firing rate are of paramount importance^{2,21}. Here we have emphasized the homeostatic aspect of differentiation, whereby steady-state changes are minimized after the transient response. Thus, the fact that synaptic depression requires the first six (or so) cycles of a tone stimulus to settle to a steady-state (Fig. 4c) correlates well with human psychophysical observations that sound localization improves markedly in the first ~60 ms of a sound stimulus²⁷ or during the first few milliseconds of binaurally time-shifted sound stimuli that follow a brief masking stimulus in the owl²⁸. Any preceding activity of the NM inputs would depress the synaptic conductance to an extent that would prevent saturation of ITD tuning curves at a given intensity. This reasoning suggests that the high spontaneous firing rates observed in NM cells (~90 spikes s⁻¹; ref. 10) shorten the time needed for accurate sound localization; in other words, the inputs to NL will always be partially depressed. In the steady-state, if the stimulus frequencies approach or surpass the 'limiting frequency'^{2,21} of a synapse (about 250 Hz for the NM–NL synapse according to Fig. 1d), a depressing synapse delivers a nearly constant net synaptic current (that is, the product of frequency and event amplitude and current amplitude) to

the postsynaptic cell. This property of a depressing synapse accounts for the relative insensitivity of the NL cell to input that codes sound intensity (NM cell firing rate) and thus ensures that the degree of input synchronization (coincidence detection) is the variable encoded by the NL cell. For this situation to occur, the kinetics of synaptic depression, which determine the limiting frequency, must be tuned to the range of natural firing frequencies occurring at the synapse. □

Methods

Characterization of NM–NL synaptic depression *in vitro*

Chicks at E18–21 were decapitated and a section of the brainstem was removed as described¹². Brain slices (200 μ m thick) were cut and perfused in a submerged recording chamber with Ringer's solution that contained (in mM): 130 NaCl, 3 KCl, 2 CaCl₂, 2 MgCl₂, 1.2 NaH₂PO₄, 26 NaHCO₃ and 10 dextrose, bubbled with 95% O₂ and 5% CO₂ (pH 7.4) at 33–35 °C. The recording chamber was mounted on a fixed stage, upright microscope.

We identified neurons by visualization on a video monitor using infrared-DIC (differential interference contrast) optics. NL neurons were recorded in current clamp using patch pipettes in the whole-cell mode that contained (in mM): 130 KCH₃SO₄, 2 MgCl₂, 5 KCl, 10 HEPES, 2 Na-ATP, 0.2 Na-GTP and 0.1 EGTA (pH 7.2). A monopolar stimulating electrode (a patch pipette filled with Ringer's solution) was placed near the soma of an ipsilateral NM neuron. We varied stimulus intensity to establish the threshold for a mono-amplitude EPSP. The stimulating electrode position was varied to ensure that the EPSP was evoked only by stimulation in a 25- μ m diameter window. To ensure that the observed EPSP depression reflected changes of synaptic function rather than a systematic change of postsynaptic input conductance, we injected trains of depolarizing currents into the soma. Depression of the current-evoked voltage responses was not seen (data not shown).

Biophysical model

Our model was implemented in the neurophysiological simulation environment

NEURON¹⁸ using 'hoc' interpreter codes, which are available from W.J.S. on request. Each presynaptic action potential triggered a postsynaptic depolarizing conductance increase that had peak conductance, G , and a time course described by

$$g(t) = Gxe^{(1-x)t} \quad (1)$$

$$\text{where } x = (t - t_{\text{onset}})/\tau_d \quad (2)$$

τ_d is the time constant governing the conductance increase, and t_{onset} is the arrival time of a presynaptic spike. The value of τ_d was set to 0.26 ms (ref. 12). If there is no synaptic depression then $G = G_{\text{max}}$, otherwise G varies with the interval since the last transmitter release according to equation (4). The resulting synaptic current, I , is given by

$$I = g(t) \times (V - E) \quad (3)$$

where V is membrane potential and E is current reversal potential (set as 0 mV).

In simulations with synaptic depression, we calculated peak synaptic conductance, G_{n+1} , for the $(n + 1)$ th synaptic event during a pure tone sound stimulus using a modified form of equation (2) from ref. 21:

$$G_{n+1} = k[G_n(1 - U)e^{-\Delta t/\tau_f} + G_{\text{max}}(1 - e^{-\Delta t/\tau_s})] + (1 - k)[G_n(1 - U)e^{-\Delta t/\tau_s} + G_{\text{max}}(1 - e^{-\Delta t/\tau_s})] \quad (4)$$

where G_{max} (9.0 nS) is the peak synaptic conductance after the synapse has experienced an 'infinite' rest; U (0.6) is the fraction of available transmitter released; τ_f (15 ms) and τ_s (1.1 s) are time constants of the fast and slow recovery processes, respectively; Δt is time elapsed since the last transmitter release, and k (0.3) is the fast fraction of the combined (fast + slow) recovery processes.

The firing pattern for each NM cell is characterized by two independent variables: average firing frequency and the jitter in time of spike initiation. This independence allowed spike frequency and jitter to be computed and varied separately using previously developed algorithms¹². Spike probability during each sound cycle was given by the ratio of average NM firing rate divided by tone frequency. At cycle onset, a random number between 0 and 1 was drawn from a uniform probability distribution. If this number was less than the spike probability ratio, a spike occurred during that cycle. NM spike jitter was specified as the standard deviation of the dispersion of NM spikes around the phase of the tone where NM cell firing occurred on average. The amount of jitter affects vector strength, an established measure of phase-locking⁶. In these simulations, jitter at each tone frequency was varied as required to produce a vector strength of 0.76, a value measured *in vivo* at a sound frequency of about 600 Hz (ref. 10). Vector strength remains constant for sound intensities >20 dB above threshold¹⁹ and was kept so for all stimulus intensities.

Equations describing the membrane conductances of our NL cell model and its detailed geometry are given in Supplementary Information. Briefly, our model NL cell had a 20- μ m diameter soma with nine ventral and nine dorsal dendrites²⁹, each of which had a uniform distribution of NM synapses around their midpoint. The model behaved like a point neuron because there was minimal attenuation or prolongation of EPSPs transmitted from dendrite to soma. A myelinated axon was connected to the soma by an initial segment to enable unambiguous counting of propagated action potentials. A 'leakage' conductance distributed over the membrane resulted in a hyperpolarizing input conductance of 72 M Ω , as measured at 37°C (ref. 12). The reversal potential of the leakage current maintained resting potential at -60 mV (ref. 12). The cell contained a Hodgkin-Huxley-like Na⁺ and K⁺ conductance. The K⁺ conductance was based on voltage-clamp measurements made at 37°C in NM neurons³⁰, with slight adjustments to match outward rectification in NL cells¹². The NL cell somatic Na⁺ conductance was formulated to replicate several NL spiking properties measured in current clamp¹².

Received 23 May; accepted 17 October 2002; doi:10.1038/nature01248.

1. Tsodyks, M., Pawelzik, K. & Markram, H. Neural networks with dynamic synapses. *Neural Comput.* **10**, 821–835 (1998).
2. Abbott, L. F., Varela, J. A., Sen, K. & Nelson, S. B. Synaptic depression and cortical gain control. *Science* **275**, 220–224 (1997).
3. Dobrunz, L. E. & Stevens, C. F. Heterogeneity of release probability, facilitation, and depletion at central synapses. *Neuron* **18**, 995–1008 (1997).
4. Nadim, F., Manor, Y., Kopell, N. & Marder, E. Synaptic depression creates a switch that controls the frequency of an oscillatory circuit. *Proc. Natl Acad. Sci. USA* **96**, 8206–8211 (1999).
5. Parks, T. N. & Rubel, E. W. Organization and development of brain stem auditory nuclei of the chicken: organization of projections from n. magnocellularis to n. laminaris. *J. Comp. Neurol.* **164**, 435–448 (1975).
6. Goldberg, J. M. & Brown, P. B. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophysiol.* **32**, 613–636 (1969).
7. Yin, T. C. & Chan, J. C. Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* **64**, 465–488 (1990).
8. Carr, C. E. & Konishi, M. A circuit for detection of interaural time differences in the brain stem of the barn owl. *J. Neurosci.* **10**, 3227–3246 (1990).
9. Zhou, N. & Parks, T. N. Pharmacology of excitatory amino acid neurotransmission in nucleus laminaris of the chick. *Hear. Res.* **52**, 195–200 (1991).
10. Warchol, M. E. & Dallos, P. Neural coding in the chick cochlear nucleus. *J. Comp. Physiol. A* **166**, 721–734 (1990).
11. Pena, J. L., Viète, S., Albeck, Y. & Konishi, M. Tolerance to sound intensity of binaural coincidence detection in the nucleus laminaris of the owl. *J. Neurosci.* **16**, 7046–7054 (1996).
12. Reyes, A. D., Rubel, E. W. & Spain, W. J. *In vitro* analysis of optimal stimuli for phase-locking and time-delayed modulation of firing in avian nucleus laminaris neurons. *J. Neurosci.* **16**, 993–1007 (1996).
13. Funabiki, K., Koyano, K. & Ohmori, H. The role of GABAergic inputs for coincidence detection in the neurons of nucleus laminaris of the chick. *J. Physiol. (Lond.)* **508**, 851–869 (1998).
14. Bruckner, S. & Hyson, R. L. Effect of GABA on the processing of interaural time differences in nucleus laminaris neurons in the chick. *Eur. J. Neurosci.* **10**, 3438–3450 (1998).

15. Yang, L., Monsivais, P. & Rubel, E. W. The superior olivary nucleus and its influence on nucleus laminaris: a source of inhibitory feedback for coincidence detection in the avian auditory brainstem. *J. Neurosci.* **19**, 2313–2325 (1999).
16. Agmon-Snir, H., Carr, C. E. & Rinzel, J. The role of dendrites in auditory coincidence detection. *Nature* **393**, 268–272 (1998).
17. Jackson, H. & Rubel, E. W. Ontogeny of behavioral responsiveness to sound in the chick embryo as indicated by electrical recordings of motility. *J. Comp. Physiol. Psychol.* **92**, 682–696 (1978).
18. Hines, M. L. & Carnevale, N. T. The NEURON simulation environment. *Neural Comput.* **9**, 1179–1209 (1997).
19. Koppl, C. Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, *Tyto alba*. *J. Neurosci.* **17**, 3312–3321 (1997).
20. Lilly, A. W. & North, K. A. K. An electrical investigation of effects of repetitive stimulation on mammalian neuromuscular junction. *J. Neurophysiol.* **16**, 509–527 (1953).
21. Tsodyks, M. V. & Markram, H. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl Acad. Sci. USA* **94**, 719–723 (1997).
22. Brenowitz, S. & Trussell, L. O. Maturation of synaptic transmission at end-bulb synapses of the cochlear nucleus. *J. Neurosci.* **21**, 9487–9498 (2001).
23. Dobrunz, L. E., Huang, E. P. & Stevens, C. F. Very short-term plasticity in hippocampal synapses. *Proc. Natl Acad. Sci. USA* **94**, 14843–14847 (1997).
24. Harata, N. *et al.* Limited numbers of recycling vesicles in small CNS nerve terminals: implications for neural signaling and vesicular cycling. *Trends Neurosci.* **24**, 637–643 (2001).
25. von Gersdorff, H., Borst, J. & Gerard, G. Short-term plasticity at the calyx of Held. *Nature Rev. Neurosci.* **3**, 55–64 (2002).
26. Jeffress, L. A. Mathematical and electrical models of auditory detection. *J. Acoust. Soc. Am.* **44**, 187–203 (1968).
27. Tobias, J. V. & Zerin, S. Lateralization thresholds as a function of stimulus duration. *J. Acoust. Soc. Am.* **31**, 1591–1594 (1959).
28. Wagner, H. A temporal window for lateralization of interaural time difference by barn owls. *J. Comp. Physiol. A* **169**, 281–289 (1991).
29. Smith, D. J. & Rubel, E. W. Organization and development of brain stem auditory nuclei of the chicken: dendritic gradients in nucleus laminaris. *J. Comp. Neurol.* **186**, 213–239 (1979).
30. Reyes, A. D., Rubel, E. W. & Spain, W. J. Membrane properties underlying the firing of neurons in the avian cochlear nucleus. *J. Neurosci.* **14**, 5352–5364 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank R. Lee for technical help and J. Simon for insight in using NEURON to model sound localization. This work was supported by a VA Merit Review and a grant from the National Institute for Deafness and Communication Disorders.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.J.S. (e-mail: spain@u.washington.edu).

Extinction-induced upregulation in AMPA receptors reduces cocaine-seeking behaviour

Michael A. Sutton*†, Eric F. Schmidt*†, Kwang-Ho Choi‡, Christina A. Schadt*, Kim Whisler‡, Diana Simmons‡, David A. Karanian*, Lisa M. Monteggia‡, Rachael L. Neve§ & David W. Self‡

* Division of Molecular Psychiatry and the Interdepartmental Neuroscience Program, Yale University School of Medicine, New Haven, Connecticut 06520, USA

‡ Department of Psychiatry, The Seay Center for Basic and Applied Research in Psychiatric Illness, The University of Texas Southwestern Medical Center, Dallas, Texas 75390-9070, USA

§ Department of Psychiatry, Harvard Medical School, McLean Hospital, Belmont, Massachusetts 02478, USA

† These authors contributed equally to this work

Cocaine addiction is thought to involve persistent neurobiological changes that facilitate relapse to drug use despite efforts to abstain. But the propensity for relapse may be reduced by extinction training—a form of inhibitory learning that progressively reduces cocaine-seeking behaviour in the absence of cocaine

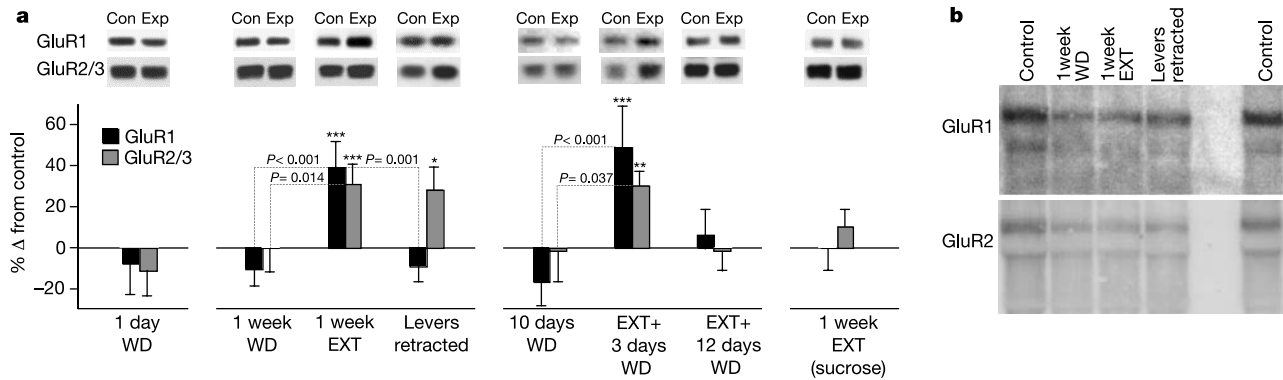


Figure 1 Extinction training increases the amount of AMPA receptor subunits in the NAC shell during cocaine withdrawal. **a**, Top, GluR1 and GluR2/3 protein in Nac shell homogenates from experimental (exp) and control (con) tissue. Bottom, protein quantities are expressed as a percentage change from group-matched untreated control tissue. Quantities of GluR1 ($F_{8,190} = 5.000$, $P < 0.001$) and GluR2/3 ($F_{8,181} = 3.231$, $P = 0.002$) increase after extinction training (EXT), as compared with cocaine withdrawal (WD) alone or when access to response levers is denied (dotted lines). Asterisks indicate

that the experimental groups ($n = 9-19$) differ significantly from pooled untreated controls for GluR1 ($100 \pm 3.0\%$, $n = 108$) and GluR2/3 ($100 \pm 3.1\%$, $n = 103$) by Fisher's test, $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$. **b**, Northern blots show decreases in mRNA for both GluR1 (50–78%) and GluR2 (56–64%) after 1 week of cocaine withdrawal (WD), with or without extinction training (EXT), as compared with untreated controls in pooled NAC shell samples.

reward¹. Here we show that extinction training during withdrawal from chronic cocaine self-administration induces experience-dependent increases in the GluR1 and GluR2/3 subunits of AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionate) glutamate receptors in the nucleus accumbens shell, a brain region that is critically involved in cocaine reward²⁻⁵. Increases in the GluR1 subunit are positively associated with the level of extinction achieved during training, suggesting that GluR1 may promote extinction of cocaine seeking. Indeed, viral-mediated overexpression of both GluR1 and GluR2 in nucleus accumbens shell neurons facilitates extinction of cocaine- but not sucrose-seeking responses. A single extinction training session, when conducted during GluR subunit overexpression, attenuates

stress-induced relapse to cocaine seeking even after GluR overexpression declines. Our findings indicate that extinction-induced plasticity in AMPA receptors may facilitate control over cocaine seeking by restoring glutamatergic tone in the nucleus accumbens, and may reduce the propensity for relapse under stressful situations in prolonged abstinence.

Rats were trained to press a lever for intravenous cocaine injections for 4 h d^{-1} for 12 d. After cocaine self-administration, rats were allowed to extinguish cocaine-seeking responses in the absence of cocaine reward for 4 h d^{-1} in the self-administration test chambers or were confined to their home cages for the same 1-week withdrawal period. In animals that experienced extinction training, amounts of the AMPA receptor subunits GluR1 and GluR2/3

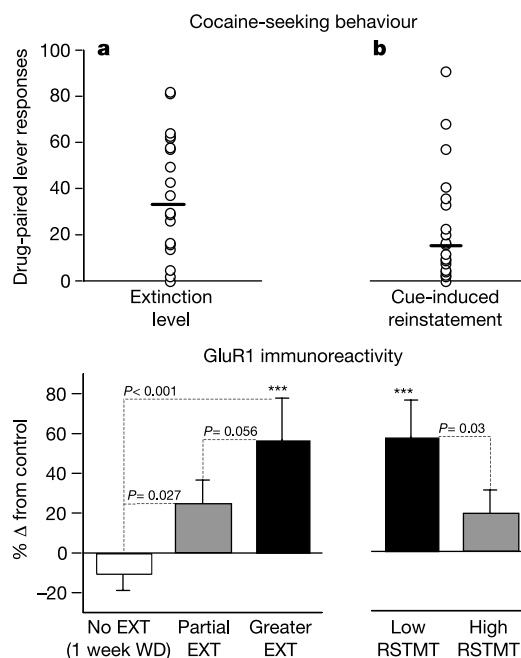


Figure 2 Increases in GluR1 are associated with extinction of cocaine-seeking and resistance to reinstatement. **a**, GluR1 in the NAc shell increases progressively from animals with no extinction training (No EXT) to animals showing partial (above median) and greater (below median) extinction of drug-paired lever responses over the last two out of six training sessions ($F_{3,85} = 7.940$, $P < 0.001$). **b**, Increases in GluR1 are associated

with resistance to cue-induced cocaine seeking in a 1-h reinstatement (RSTMT) test ($F_{2,59} = 9.723$, $P < 0.001$). Black bars indicate the median cocaine-seeking response rate of 19 animals. Baseline responses for 1 h before cue presentation averaged less than ten in both low and high RSTMT groups. Experimental groups differ from pooled untreated controls by Fisher's test, $***P < 0.001$.

increased by 39% and 31% in the nucleus accumbens (NAc) shell (but not core) region, relative to both untreated controls and cocaine-trained animals that remained in the home cages during withdrawal (Fig. 1a). Extinction-induced increases in GluR1 and GluR2/3 persisted for at least 3 d after cessation of extinction training but declined to control amounts after 12 d. By contrast, amounts of GluR1 and GluR2/3 were similar to control values after 1, 7 or 10 d of cocaine withdrawal in the home cages, although messenger RNAs for GluR1 and GluR2 actually decreased by 22–50% after 7 d, regardless of extinction training experience (Fig. 1b). Thus, extinction-induced plasticity does not involve the regulation of GluR1 and GluR2 gene expression.

Extinction-induced increases in GluR1 and GluR2/3 were not attributable to extinction-related learning or other procedural influences, because they did not increase in sucrose-trained animals that extinguished sucrose-seeking responses to a similar degree under identical extinction training procedures (Fig. 1a). Amounts of GluR1 also failed to increase in cocaine-trained animals that were prevented from responding (levers retracted) but that experienced similar daily excursions to the self-administration test chambers (Fig. 1a), suggesting that increases in GluR1 require extinction of cocaine-seeking responses. Indeed, GluR1 increases were associated positively with the level of extinction achieved during training (Fig. 2a). Thus, GluR1 increased by 58% in animals whose cocaine-seeking responses fell below the median in the final two extinction sessions, whereas GluR1 increased by only a marginal 24% (approaching significance, $P = 0.057$) in partially extinguished animals responding above the median, although both subgroups showed increases relative to self-administering animals that did not receive extinction training.

Conversely, the ability of cocaine-specific 'cues' to reinstate cocaine-seeking after extinction was negatively associated with increases in GluR1 (Fig. 2b). Animals showing relatively no cue response (below the median) had greater increases in GluR1 than had animals responsive to the cues (above the median). By contrast, increases in GluR2/3 were not related to extinction or cue responsiveness ($P > 0.1$), and these subunits increased even without access to the levers (Fig. 1a), indicating that repeated exposure to the cocaine-associated test chambers, and not extinction of cocaine-seeking responses, is sufficient to increase GluR2/3.

Given that extinction-induced increases in GluR1 were associated with a reduction in cocaine-seeking during both extinction and cue reinstatement, we examined whether this experience-dependent plasticity would act reciprocally to facilitate extinction from cocaine self-administration. This hypothesis was tested by transiently overexpressing GluR1 and GluR2 subunits in NAc shell neurons before extinction testing. Figure 3 shows effective *in vivo* transfection of a bacterial reporter gene (*lacZ*) and GluR1 after NAc shell infusions of herpes simplex virus (HSV) vectors encoding these genes. The right side infused with HSV-*lacZ* showed prominent β -galactosidase activity (Fig. 3a), whereas the left side infused with HSV-GluR1 showed a strong increase in GluR1 immunoreactivity (Fig. 3b) in consecutive immunostained brain sections.

HSV-mediated overexpression was confined to the medial NAc shell in both stained sections and immunoblots of NAc shell samples probed with antibodies against β -galactosidase (Fig. 3d) as compared with adjacent core tissue. Higher magnification showed that GluR1 overexpression occurred throughout cell bodies and processes (Fig. 3b, inset), indicating its transport to dendritic sites of action. Previous work found that HSV-GluR1 overexpression is neuron-specific and facilitates AMPA-mediated physiological responses, and a similar overexpression pattern in NAc has been shown with HSV-GluR2 (refs 6, 7). HSV-mediated overexpression was undetectable after 6 d in both stained brain sections and immunoblots of NAc shell tissue (Fig. 3a, b, d), and showed no evidence for toxicity or gliosis in Nissl-stained sections and immunoblots of glial fibrillary acidic protein (GFAP; Fig. 3c, e).

HSV-mediated overexpression of either GluR1 or GluR2 in the NAc shell reduced cocaine-seeking responses during extinction as compared with sham-operated (cannulated) and HSV-*lacZ*-infused controls (Fig. 4a). The effect was most prominent during the initial extinction training session when response rates were maximal, but overexpression of GluR1 and GluR2 also shortened the latency to achieve extinction criterion by 1.5–2 training sessions over 5 d of testing. Possible performance and/or cognitive deficits were negated by the fact that animals overexpressing GluR1 and GluR2 showed within- and between-session declines that were typical of extinction, and similar HSV infusions failed to alter extinction responding in sucrose-trained animals (Fig. 4b). Thus, both AMPA subunits selectively facilitate extinction from cocaine self-administration, possibly by directly reducing the motivation to seek cocaine, or indirectly by promoting cocaine-specific extinction learning mechanisms.

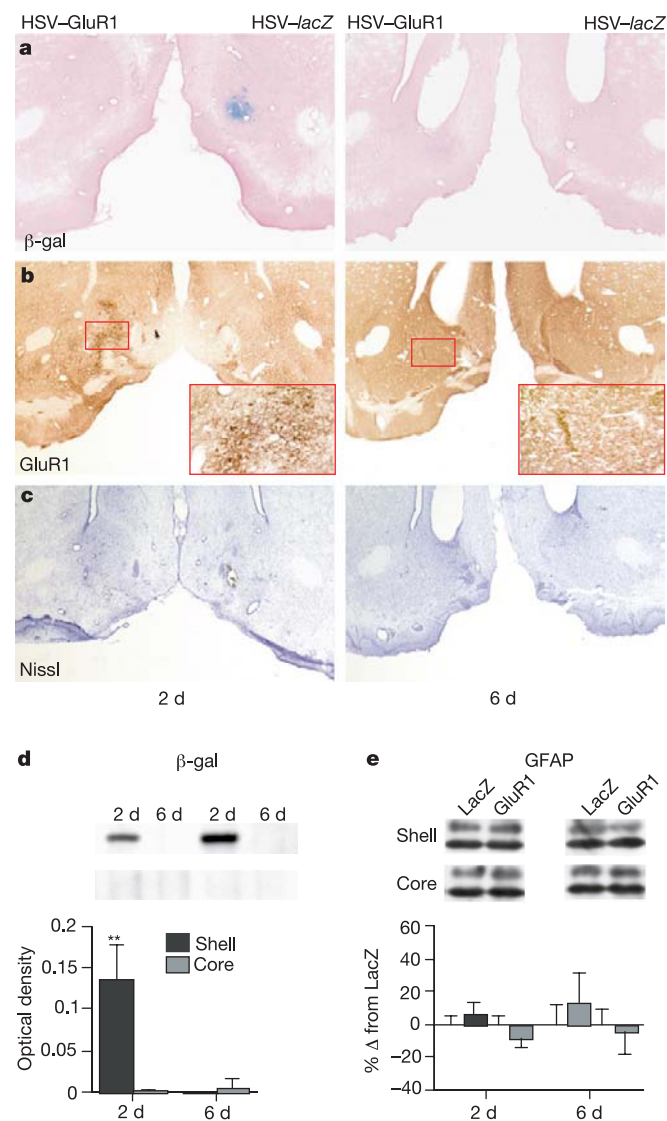


Figure 3 Transient viral-mediated overexpression of β -galactosidase and GluR1 in NAc *in vivo*. Shown is β -galactosidase activity (**a**) and GluR1 immunoreactivity (**b**; inset, original magnification, $\times 200$) after intra-NAc infusions of HSV-*lacZ* and HSV-GluR1 ($2.0 \mu\text{l}$ per side; 4×10^7 infectious units per ml). Overexpression is confined to the NAc shell at 2 d and is undetectable after 6 d in either immunostained brain sections or immunoblots of NAc shell homogenates labelled with antibodies against β -galactosidase (**d**). There is no evidence for reactive gliosis in HSV-infused tissue in Nissl-stained sections (**c**) or immunoblots labelled with antibodies against GFAP (**e**). β -Galactosidase at 2 d after infusion ($n = 11$) differs from that at 6 d ($n = 6$) by *t*-test, $**P < 0.01$.

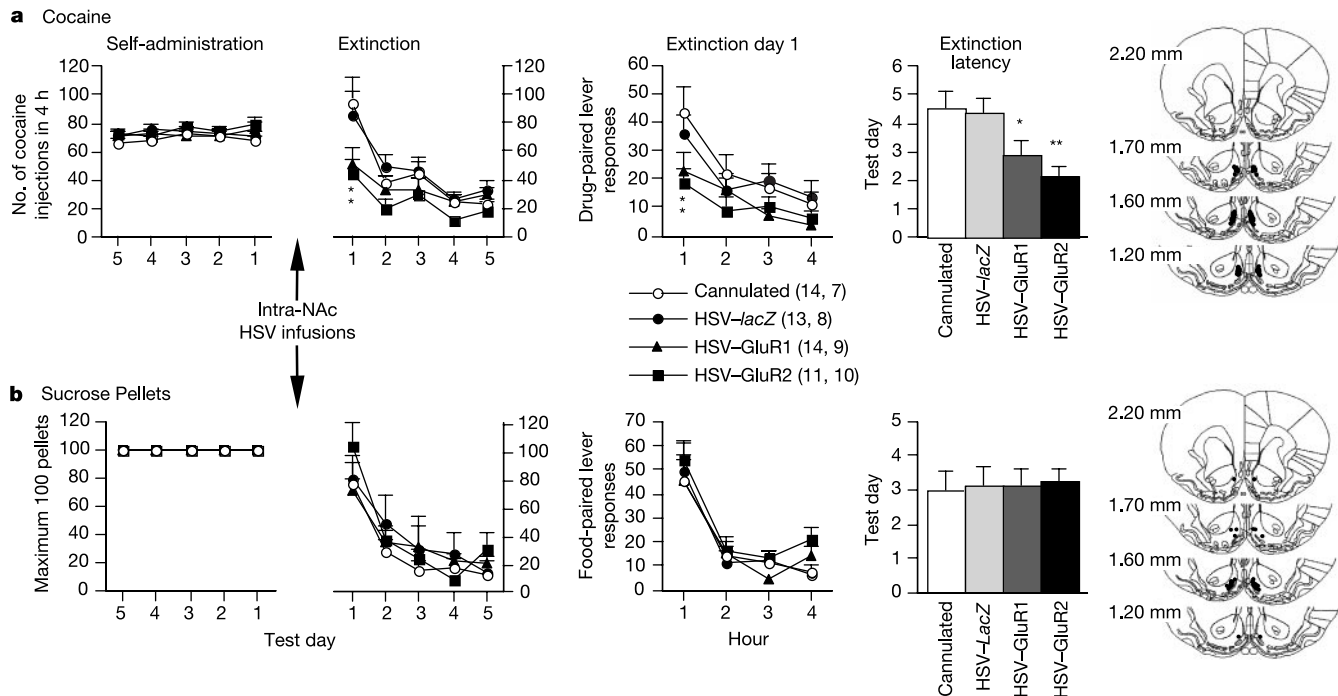


Figure 4 Effects of HSV-mediated overexpression of GluR1 and GluR2 on extinction of cocaine and sucrose seeking. **a**, Overexpression of either GluR1 or GluR2 in the NAC shell reduces cocaine-seeking responses in extinction from 15 to 20 d of cocaine self-administration ($F_{2,49} = 3.322$, $P = 0.044$; group $\times$ test day interaction, $F_{8,196} = 2.406$, $P = 0.017$), and shortens the number of test days to reach extinction criterion (<20 responses in 4 h) as compared with sham (cannulated) and HSV-*lacZ*-infused controls

($F_{2,49} = 7.397$, $P = 0.002$). Both GluR1 and GluR2 reduce high rates of cocaine seeking during the first hour of the initial extinction test ($F_{2,49} = 3.671$, $P = 0.033$). **b**, Similar intra-NAC infusions of HSV-GluR1 and HSV-GluR2 fail to alter extinction responses in animals trained to self-administer sucrose pellets. HSV infusion sites are indicated in coronal planes (1.2–2.2 mm anterior to bregma). Asterisks indicate values differ from pooled sham and HSV-*lacZ*-treated controls by Fisher's test, $*P < 0.05$ and $**P < 0.01$.

To distinguish between these possibilities, cocaine-trained animals were allowed to experience a single 4-h extinction training session during HSV-mediated overexpression of GluR1 or GluR2, and were then subjected to the remainder of extinction testing 2 weeks later (Fig. 5), at least 1 week after overexpression declined. In these animals, GluR1 and GluR2 overexpression reduced initial cocaine-seeking responses as before, but extinction rates were similar to HSV-*lacZ*-infused controls after overexpression declined, indicating that GluR1 and GluR2 do not facilitate extinction learning. Instead, these results suggest that GluR1 and GluR2 reduce the motivation for cocaine, possibly by directly facilitating excitatory transmission in the NAC shell emanating from cortical, hippocampal, amygdala and/or thalamic inputs. By contrast,

other studies have found that generalized activation of NAC neurons through exogenous AMPA infusions actually triggers cocaine seeking^{8,9}, suggesting that temporal and/or spatial integration of endogenous glutamatergic signals, or tonic overexpression, is important for these GluR1 and GluR2 effects.

Given that the extinction-induced increases in endogenous AMPA subunits were relatively transient (Fig. 1a), we tested whether such transient overexpression during extinction would affect relapse to cocaine seeking at later withdrawal times, when GluR1 and GluR2 levels return to normal. Relapse (reinstatement) was induced by exposing animals to cocaine-associated cues, brief intermittent footshock stress and intravenous priming injections of cocaine itself, because similar stimuli trigger cocaine craving in

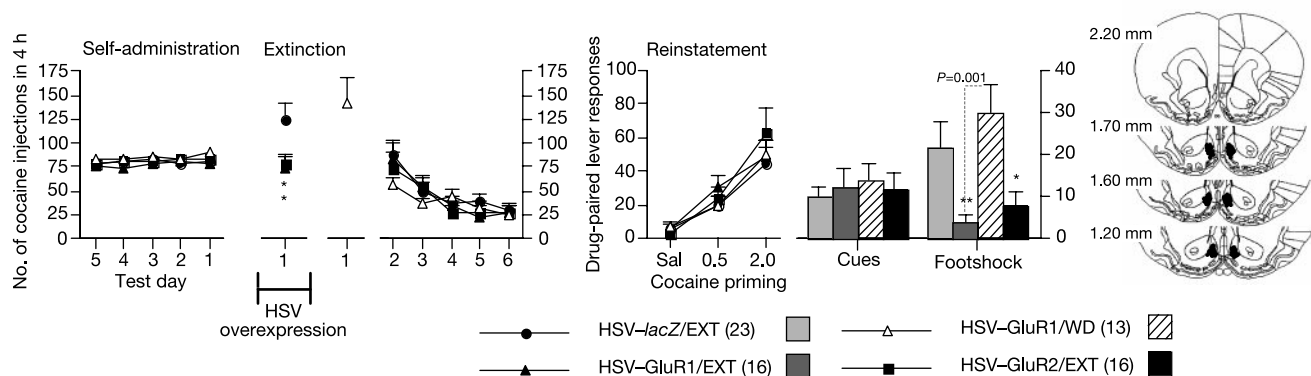


Figure 5 Prolonged effects of extinction training concomitant with overexpression of GluR1 or GluR2 on stress-induced reinstatement. A single extinction training session (2 d after HSV infusion) attenuates cocaine seeking induced by footshock stress long after overexpression declines (27 d after HSV, $F_{3,64} = 5.253$, $P = 0.003$). Cocaine seeking induced by cue presentation and cocaine priming injections is not altered. Temporal

dissociation of GluR1 overexpression from extinction training fails to attenuate cocaine seeking induced by footshock stress (HSV-GluR1/WD). HSV infusion sites are indicated in coronal planes (1.2–2.2 mm anterior to bregma). Asterisks indicate values differ from HSV-*lacZ* by Fisher's test, $*P < 0.05$ and $**P < 0.01$.

humans^{10–12}. Unexpectedly, prior overexpression of GluR1 and GluR2, when concomitant with a single extinction training session, severely attenuated the subsequent ability of footshock stress to trigger cocaine-seeking responses (Fig. 5). Stress-induced relapse was attenuated despite the fact that neither extinction rates nor cue- and drug-induced reinstatement was altered in these animals and that testing was conducted 3 weeks after overexpression of GluR1 and GluR2 declined. Thus, transient increases in endogenous GluR1 and GluR2/3 induced during extinction training could profoundly attenuate stress-induced relapse, despite the normalization of AMPA subunits days after cessation of extinction training. These results also suggest that stress induces relapse through neural pathways that differ from those used by cues and drugs.

Because experience-dependent plasticity in endogenous GluR1 involved a specific interaction with cocaine-seeking responses (Figs 1a and 2), we tested whether transient GluR1 overexpression requires concurrent extinction training experience to attenuate subsequent stress-induced cocaine seeking. In this group, HSV-GluR1 was infused at the same time of cocaine withdrawal as in the other HSV-treated groups, but the initial extinction training session was delayed until 2 weeks after overexpression declined (Fig. 5). Notably, temporal dissociation of GluR1 overexpression from extinction training completely failed to attenuate subsequent stress-induced relapse. Together, these results suggest that extinction-induced increases in GluR1 in the NAc shell form a reciprocal interaction with extinction experience itself that subsequently weakens neural substrates of stress-induced relapse.

Repeated, passive administration of cocaine reduces AMPA-mediated excitability in the NAc shell, but not core, for at least 14 d of withdrawal^{13,14}. This reduction in excitability involves long-term depression in glutamatergic synapses emanating from ventral areas of the medial prefrontal cortex¹⁴. Reduced excitability correlates with behavioural sensitization to cocaine and amphetamine, and corresponds to a reduction in both GluR1 and GluR2 mRNA and protein^{15,16}, although increases in GluR1 have been detected at later times in sensitized animals¹⁷. Our study found that extinction from chronic cocaine self-administration produced marked increases in GluR1 and GluR2/3 (recent work suggests that GluR2 is increased by extinction training) in the NAc shell, despite a substantial decrease in mRNA levels, suggesting that synaptic regulation may underlie the increases in AMPA subunits.

It is possible that extinction-related activity in cortical or other glutamatergic inputs to the NAc shell facilitates trafficking and stabilization of GluR1 and GluR2 in synaptic membranes, similar to experience- and activity-dependent glutamate receptor trafficking in other brain regions^{18,19}. Indeed, most NAc neurons are highly activated during extinction from cocaine self-administration, and this activity persists or increases even after cocaine-seeking responses diminish²⁰. In addition, differential regulation of GluR1, but not GluR2/3, with and without lever access might reflect subunit-specific synaptic trafficking mechanisms. For example, GluR1 trafficking is induced by synaptic activity and phosphorylation, whereas GluR2 is cycled continuously, regardless of synaptic activity^{21,22}. Alternatively, increases in AMPA subunits might represent a compensatory response to reduced extracellular glutamate during cocaine withdrawal, because this reduction is exacerbated by context-related factors^{23,24}. But AMPA subunits are increased by extinction conditions, and not by cocaine withdrawal alone, and AMPA receptors fail to upregulate with cortical denervation²⁵.

In any event, given that chronic cocaine use is associated with prolonged hypoactivity in prefrontal cortical areas^{26–28}, extinction training might restore cortical-accumbal tone by upregulating AMPA receptor responses. Our results also indicate that such upregulation may benefit cocaine addicts both directly, by facilitating control over the urge for cocaine, and indirectly, by attenuating stress-induced cravings in prolonged abstinence. In addition, they suggest that behaviour-based approaches have the potential to

reverse or ameliorate the harmful neurobiological and behavioural consequences of chronic drug use. □

Methods

Behavioural procedures

Male Sprague–Dawley rats (300–350 g) were housed individually in accordance with NIH standards and received lever-press training for sucrose pellets (45 mg) before intravenous catheter implantation²⁹. Age- and group-matched biochemical controls also received lever-press training but remained in their home cages until tissue collection. A subset of catheterized animals was implanted with 26 gauge bilateral guide cannulae in the NAc shell for HSV infusions (+1.7 mm anterior, ±0.8 mm lateral to bregma, and –5.7 mm from dura). HSV vectors were infused 1 mm below the guide cannulae through 33 gauge injectors over 10 min (2 µl per side). Animals self-administered cocaine injections by performing a single lever press (fixed ratio 1) in daily 4-h test sessions (5–6 d week^{–1}) in two-lever operant test chambers as described²⁹. A compound stimulus (cue light plus pump noise plus house lights off) accompanied each 5-s injection of cocaine during the 15-s time-out period that followed the onset of each injection. Average cocaine intake over the last 5 d of self-administration ranged from 38.6 to 43.6 mg per kg (body weight) per d (1.0 mg per kg per 10-s injection) for groups in AMPA biochemistry and from 34.6 to 42.5 mg per kg per d (0.5 mg per kg per 5-s injection) for groups in HSV experiments. Sucrose-trained animals were maintained at 85% initial body weight and self-administered sucrose pellets (fixed ratio 1) for 4 h d^{–1} or until a maximum of 100 pellets was earned.

Extinction training was conducted in daily 4-h sessions in the absence of cocaine or cue reinforcement. Both sucrose- (fed *ad libitum*) and cocaine- (1-week EXT) trained animals extinguished to a mean 7.5 ± 3.0 responses per h and 7.0 ± 1.4 responses per h, before biochemical analysis of AMPA subunits. Reinstatement sessions were identical to extinction sessions, except that reinstating stimuli were presented before or during the last hour according to the following daily sequence: cue presentation (every 2 min for 1 h), after an intravenous cocaine priming injection (saline, 0.5 and 2.0 mg per kg, in counterbalanced order), after 30 min of intermittent (random interval averaging 30 s) footshock stress (1.0 mA for 0.5 s). Inactive lever responding in reinstatement tests with cues, cocaine, footshock stress were similar to saline priming ($F_{4,256} = 1.311$, $P = 0.266$).

Immunoblot procedures

Experimental rats alternating with age- and group-matched biochemical controls were removed from their home cages and immediately decapitated in a separate room 12–16 h, 3 or 12 d after the last extinction training session. NAc shell and core tissue samples were collected and analysed by immunoblotting as described²⁹. Blots were immunolabelled with polyclonal rabbit antibodies against GluR1 and GluR2/3 (1:5,000), and antibodies against β-galactosidase (1:10,000) from Chemicon, and monoclonal mouse antibody against GFAP (Sigma; 1:20,000), followed by secondary labelling with polyclonal peroxidase-conjugated antibodies (goat antibodies against rabbit or mouse IgG, Chemicon; 1:2,000). We detected immunoreactivity using enhanced chemiluminescence (NEN) and quantified it using NIH Image 1.57, as described²⁹.

Histochemical staining

HSV-infused animals were perfused with 4% paraformaldehyde in PBS, post-fixed at 4 °C in 4% paraformaldehyde and cryoprotected in 30% sucrose for at least 3 d. Consecutive 35-µm coronal brain sections were stored in PBS (for X-Gal staining) or 0.1% sodium azide in PBS (for immunostaining). For X-Gal staining, we washed free-floating sections in 10 mM PBS and placed them in a reaction buffer (3.1 mM potassium ferricyanide, 3.1 mM potassium ferrocyanide, 150 mM NaCl, 1 mM MgCl₂, 0.01 M sodium deoxycholate, 0.02% Nonidet P-40, 0.2 mg ml^{–1} X-Gal in 50 mM phosphate buffer) at 37 °C for 2–4 h until the signal was optimal. Sections were then washed, mounted and counterstained with Nuclear Fast Red (Vector Labs). For GluR1 staining, we treated the washed sections with 1% H₂O₂ in PBS for 30 min, and blocked them with 2.5% normal goat serum (NGS) in PBS containing 0.3% Triton X-100 for 1 h at 23 °C. Sections were incubated overnight with antibodies against GluR1 (1:1,000) in 1% NGS in PBS plus 0.3% Triton X-100, then for 1 h with biotinylated secondary goat antibodies against rabbit IgG (1:200) in 1% NGS in PBS, and finally for 1 h with avidin–biotin–peroxidase complex (Vector Labs). Peroxidase activity was labelled for 10 min with a 3,3'-diaminobenzidine kit (Pierce).

Northern blot procedures

Complementary DNAs for GluR1, GluR2 and β-actin were obtained by polymerase chain reaction amplification as described³⁰ and cloned into pCRII-TOPO (Invitrogen). DNA templates were linearized with *Xho*I (GluR1) or *Xba*I (GluR2, β-actin) before transcription with antisense probes. *In vitro* transcription was carried out with SP6 polymerase (Roche) in the presence of [³²P]labelled UTP (>800 Ci mmol^{–1}; NEN), with excess nucleotides removed by two ethanol precipitations.

We extracted total RNA by homogenization in Trizol (Invitrogen) from pooled NAc shell samples (5–11 per group) from self-administering animals for comparison with two separate controls (5 and 11 per group) using contralateral tissue from immunoblot experiments (left and right counterbalanced). Northern analysis was done with the NorthernMax–Glyoxal system (Ambion). In brief, denatured total RNA samples (28 µg) were fractionated by 1% agarose/glyoxal gel electrophoresis, stained in 50 µg ml^{–1} ethidium bromide and transferred to nylon membranes. Ultraviolet-crosslinked nylon membranes were prehybridized for 30 min at 68 °C in ULTRAhyb solution (Ambion), and hybridized overnight with 10⁶ c.p.m. per ml of probe at 68 °C. Blots were washed once in 2 × SSC for 10 min (23 °C) and twice in 0.1 × SSC for 15 min (68 °C). GluR1 blots were exposed in PhosphorImager cassettes (Molecular Devices) stripped and reprobed with

GluR2 and then again with β -actin. Levels of mRNA were densitized using ImageQuant (Molecular Devices). Blots for GluR1 and GluR2 revealed characteristic RNA species³⁰, and the main bands for GluR1 (5.2 kilobases) and GluR2 (5.9 and 3.9 kilobases) were quantified and normalized relative to β -actin.

Received 16 May; accepted 11 October 2002; doi:10.1038/nature01249.

1. Bouton, M. E. & Schwartzberg, D. Sources of relapse after extinction in pavlovian and instrumental learning. *Clin. Psych. Rev.* **11**, 123–140 (1991).
2. Roberts, D. C. S. & Koob, G. F. Extinction and recovery of cocaine self-administration following 6-hydroxydopamine lesions of the nucleus accumbens. *Pharmacol. Biochem. Behav.* **12**, 781–787 (1980).
3. Pettit, H. O., Ettenberg, A., Bloom, F. E. & Koob, G. F. Destruction of dopamine in the nucleus accumbens selectively attenuates cocaine but not heroin self-administration in rats. *Psychopharmacol.* **84**, 167–173 (1984).
4. Zito, K. A., Vickers, G. & Roberts, D. C. S. Disruption of cocaine and heroin self-administration following kainic acid lesions of the nucleus accumbens. *Pharmacol. Biochem. Behav.* **23**, 1029–1036 (1985).
5. Carlezon, W. A. Jr, Devine, D. P. & Wise, R. A. Habit-forming actions of nomifensine in nucleus accumbens. *Psychopharmacol.* **122**, 194–197 (1995).
6. Neve, R. L., Howe, J. R., Hong, S. & Kalb, R. G. Introduction of the glutamate receptor subunit 1 into motor neurons in vitro and in vivo using a recombinant herpes simplex virus. *Neuroscience* **79**, 435–447 (1997).
7. Kelz, M. B. et al. Expression of the transcription factor Δ FosB in the brain controls sensitivity to cocaine. *Nature* **401**, 272–276 (1999).
8. Cornish, J. L., Duffy, P. & Kalivas, P. W. A role for nucleus accumbens glutamate transmission in the relapse to cocaine-seeking behavior. *Neuroscience* **93**, 1359–1367 (1999).
9. Cornish, J. & Kalivas, P. W. Glutamate transmission in the nucleus accumbens mediates relapse in cocaine addiction. *J. Neurosci.* **20**, RC89 (2000).
10. Jaffe, J. H., Cascella, N. G., Kumor, K. M. & Sherer, M. A. Cocaine-induced cocaine craving. *Psychopharmacology* **97**, 59–64 (1989).
11. Robbins, S. J., Ehrman, R. N., Childress, A. R. & O'Brien, C. P. Relationships among physiological and self-report responses produced by cocaine-related cues. *Addict. Behav.* **22**, 157–167 (1997).
12. Sinha, R., Catapano, D. & O'Malley, S. Stress induced craving and stress response in cocaine dependent individuals. *Psychopharmacology* **142**, 343–351 (1999).
13. White, F. J., Hu, X. T., Zhang, X. F. & Wolf, M. E. Repeated administration of cocaine or amphetamine alters neuronal responses to glutamate in the mesoaccumbens dopamine system. *J. Pharmacol. Exp. Ther.* **273**, 445–454 (1995).
14. Thomas, M., Beurrier, C., Bonci, A. & Malenka, R. Long-term depression in the nucleus accumbens: a neural correlate of behavioral sensitization to cocaine. *Nature Neurosci.* **4**, 1217–1223 (2001).
15. Lu, W., Chen, H., Xue, C. J. & Wolf, M. E. Repeated amphetamine administration alters the expression of mRNA for AMPA receptor subunits in rat nucleus accumbens and prefrontal cortex. *Synapse* **26**, 269–280 (1997).
16. Lu, W. & Wolf, M. E. Repeated amphetamine administration alters AMPA receptor subunit expression in rat nucleus accumbens and medial prefrontal cortex. *Synapse* **32**, 119–131 (1999).
17. Churchill, L., Swanson, C. J., Urbina, M. & Kalivas, P. W. Repeated cocaine alters glutamate receptor subunit levels in the nucleus accumbens and ventral tegmental area of rats that develop behavioral sensitization. *J. Neurochem.* **72**, 2397–2403 (1999).
18. Quinlan, E. M., Philpot, B. D., Huganir, R. L. & Bear, M. F. Rapid, experience-dependent expression of synaptic NMDA receptors in visual cortex in vivo. *Nature Neurosci.* **2**, 352–357 (1999).
19. Heynen, A. J., Quinlan, E. M., Bae, D. C. & Bear, M. F. Bidirectional, activity-dependent regulation of glutamate receptors in the adult hippocampus in vivo. *Neuron* **28**, 527–536 (2000).
20. Fabbriatore, A., Uzuiak, A., West, M. & Peoples, L. Comparisons of firing rates of rat nucleus accumbens neurons during cocaine self-administration and extinction. *Soc. Neurosci. Abstr.* **24**, 1736 (1998).
21. Hayashi, Y. et al. Driving AMPA receptors into synapses by LTP and CaMKII: requirement for GluR1 and PDZ domain interaction. *Science* **287**, 2262–2267 (2000).
22. Shi, S.-H., Hayashi, Y., Esteban, J. A. & Malinow, R. Subunit-specific rules governing AMPA receptor trafficking to synapses in hippocampal pyramidal neurons. *Cell* **105**, 331–343 (2001).
23. Keys, A. S., Mark, G. P., Emre, N. & Meshul, C. K. Reduced glutamate immunolabeling in the nucleus accumbens following extended withdrawal from self-administered cocaine. *Synapse* **30**, 393–401 (1998).
24. Bell, K., Duffy, P. & Kalivas, P. W. Context-specific enhancement of glutamate transmission by cocaine. *Neuropsychopharmacology* **23**, 335–344 (2000).
25. McLennan, H. The effect of deafferentation on the excitatory amino acid sensitivity of striatal neurones. *Neurosci. Lett.* **18**, 313–316 (1980).
26. Volkow, N. D. et al. Long-term frontal brain metabolic changes in cocaine abusers. *Synapse* **11**, 184–190 (1992).
27. London, E. D., Bonson, K. R., Ernst, M. & Grant, S. Brain imaging studies of cocaine abuse: implications for medication development. *Crit. Rev. Neurobiol.* **13**, 227–242 (1999).
28. Volkow, N. D. & Fowler, J. S. Addiction, a disease of compulsion and drive: involvement of the orbitofrontal cortex. *Cereb. Cortex* **10**, 318–325 (2000).
29. Schmidt, E. F. et al. Extinction training regulates tyrosine hydroxylase during withdrawal from cocaine self-administration. *J. Neurosci.* **21**, RC137 (2001).
30. Guitart, X. et al. Regulation of ionotropic glutamate receptor subunits in different rat brain areas by a preferential σ_1 receptor ligand and potential atypical antipsychotic. *Neuropsychopharmacol.* **23**, 539–546 (2000).

Acknowledgements This work was supported by US Public Health Service grants (NIDA), a postdoctoral National Research Service Award (to C.A.S.) and the Lydia Bryant Test Professorship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W.S. (e-mail: david.self@utsouthwestern.edu).

Ectopic β -chain of ATP synthase is an apolipoprotein A-I receptor in hepatic HDL endocytosis

Laurent O. Martinez*, Sébastien Jacquet*, Jean-Pierre Esteve†, Corinne Rolland*, Elena Cabezon‡, Eric Champagne§, Thierry Pineau||, Valérie Georgeaud*, John E. Walker‡, François Tercé*, Xavier Collet*, Bertrand Perret* & Ronald Barbaras*

* Institut Fédératif de Recherche Claude de Preval, IFR 30, Institut National de la Santé et de la Recherche Médicale, Unité 563, Département Lipoprotéines, et Médiateurs Lipidiques and § Département Immunologie Moléculaire et Biologie du Lymphocyte T, 31059, Toulouse cedex, France

† Institut Fédératif de Recherche Louis Bugnard, IFR 31, Institut National de la Santé et de la Recherche Médicale, Unité 531, Biologie et Pathologie Digestive, Hôpital Rangueil, 31403, Toulouse cedex, France

‡ Medical Research Council Dunn Human Nutrition Unit, Hills Road, Cambridge CB2 2XY, UK

|| Institut National de la Recherche Agronomique, Laboratoire de Pharmacologie et Toxicologie, 31931, Toulouse cedex 9, France

The effect of high-density lipoprotein (HDL) in protecting against atherosclerosis is usually attributed to its role in 'reverse cholesterol transport'¹. In this process, HDL particles mediate the efflux and the transport of cholesterol from peripheral cells to the liver for further metabolism and bile excretion. Thus, cell-surface receptors for HDL on hepatocytes are chief partners in the regulation of cholesterol homeostasis². A high-affinity HDL receptor for apolipoprotein A-I (apoA-I) was previously identified on the surface of hepatocytes^{3,4}. Here we show that this receptor is identical to the β -chain of ATP synthase, a principal protein complex of the mitochondrial inner membrane. Different experimental approaches confirm this ectopic localization of components of the ATP synthase complex and the presence of ATP hydrolase activity at the hepatocyte cell surface. Receptor stimulation by apoA-I triggers the endocytosis of *holo*-HDL particles (protein plus lipid) by a mechanism that depends strictly on the generation of ADP. We confirm this effect on endocytosis in perfused rat liver *ex vivo* by using a specific inhibitor of ATP synthase. Thus, membrane-bound ATP synthase has a previously unsuspected role in modulating the concentrations of extracellular ADP and is regulated by a principal plasma apolipoprotein.

We previously showed that apoA-I that is not associated with lipids (hereafter called free apoA-I) interacts specifically with high-affinity HDL receptors (10^{-9} M)^{3,4}, thereby representing a possible ligand for the affinity purification of HDL receptors. The passage of solubilized porcine liver plasma membrane proteins over immobilized free apoA-I (ref. 4) in surface plasmon resonance (Biacore) experiments indicated that this interaction was conserved (dissociation constant, $K_d \approx 10^{-9}$ M; Fig. 1b, sensogram 1). Subsequently, several rounds of binding and desorption allowed the purification and identification of a protein with a relative molecular mass of 50,000 (M_r 50K; Fig. 1a, lane 3). Higher amounts of apoA-I-bound proteins were recovered by affinity chromatography (using immobilized free apoA-I; Fig. 1a, lane 2) and showed a fourfold increase in binding to free apoA-I immobilized on a sensor chip (Fig. 1b, sensogram 2). We removed the 50K band from the gel and microsequenced 50 pmol of protein by protease digestion, high-performance liquid chromatography (HPLC) separation and Edman analysis. Unexpectedly, a peptide sequence derived from this protein was identical to a segment of the human β -chain of ATP synthase.

Mitochondrial ATP synthase has two main domains, F_1 and F_0 (ref. 5). The β -chain belongs to F_1 , a peripheral membrane protein

complex containing binding sites for ATP and ADP, and the catalytic site for ATP synthesis or hydrolysis. F_1 is bound on the membrane by its interaction with F_0 , an integral membrane protein complex in mammalian mitochondria that contains a transmembrane channel for protons⁵⁻⁷.

The β -chain of ATP synthase is found mostly at the inner side of the mitochondrial membrane⁵, although studies have reported its presence at the cell surface⁸⁻¹⁰. Immunofluorescence microscopy with an antibody against the β -chain or with an isotypic IgG2a confirmed the presence of the ATP synthase β -chain on the surface of immortalized human hepatocytes (IHH cells¹¹, Fig. 2a, g), but not on the surface of Chinese hamster ovary (CHO) cells (Fig. 2f, i), which do not show apoA-I high-affinity sites (unpublished data). Notably, another principal protein of ATP synthase, the α -chain, was also detected on the hepatocyte cell surface (Fig. 2d). As a control, a typical mitochondrial protein, the subunit I of cytochrome oxidase (COX-I), was not detected on the cell surface (Fig. 2e), but was visualized after permeabilization of the cells (Fig. 2h). After preincubating the cells with apoA-I, confocal immunofluorescence microscopy with an antibody against apoA-I identified a specific cell-surface signal (Fig. 2b) that was superimposed (Fig. 2c) on that of the β -chain of ATP synthase (Fig. 2a). This confirmed colocalization of the β -chain of ATP synthase and apoA-I binding at the hepatocyte cell surface.

Fluorescence-assisted flow cytometry experiments on intact cells (selected as cells excluding propidium iodide) confirmed the cell-surface localization of the β -chain specifically on HepG2 (Fig. 3a, curve 5) but not CHO (Fig. 3c, curve 5) cells. The selectivity of the response to the β -chain antibody was clearly shown by the much lower signal obtained with isotypic immunoglobulin- γ (IgG), an excess of purified F_1 -ATPase (as a competitor) or an antibody against COX-I (Fig. 3, curves 1-3, respectively). The presence of

the α -chain at the cell surface was confirmed using a specific antibody (data not shown), indicating that the whole F_1 -ATPase domain might be present on the cell surface. Incubation of hepatocytes with an excess of free apoA-I reduced the immunoreactivity of the β -chain antibody over threefold, confirming the cell-surface interaction of free apoA-I with the β -chain of ATP synthase (Fig. 3a and b, curve 4).

The association between apoA-I and the β -chain was verified by two types of competition experiment (ref. 3 and Supplementary Fig. 1): first, the β -chain antibody (Supplementary Fig. 1a) and the corresponding Fab fragments (data not shown) completely inhibited the binding of ¹²⁵I-labelled free apoA-I, as well as unlabelled ligand; second, the same antibody or Fab fragments inhibited the binding of ¹²⁵I-labelled HDL₃ (the most abundant subfraction of HDL, which binds to the high- and low-affinity binding sites through apoA-I; ref. 3) on inverted purified mitochondria (ref. 12 and Supplementary Fig. 1b), in which the F_1 fraction of ATP synthase is exposed on the outside and so avoids interference with the cell-surface low-affinity HDL-binding sites.

The presence of ATP synthase at the cell surface of lymphocytes⁸ and human endothelial cells^{9,10,13} has been reported. In lymphocytes, the interaction of ATP synthase with angiostatin suggested that it might have a role in angiogenesis. In addition, the α - and β -chains of ATP synthase have been identified as a receptor for apoE-enriched HDL^{14,15}, although the presence of this protein on the cell surface was not demonstrated nor was a role proposed in those studies. Our data clearly show the presence of the β -chain, probably as a F_1 complex, on the cell surface of HepG2, IHH cells and primary human hepatocytes¹⁶, but not on epithelial CHO cells⁸.

We measured the functional activity of cell-surface ATP synthase by adding either ADP plus ³²P, or [α -³²P]ATP, to HepG2 cells, and identifying and quantifying the nucleotides generated in culture

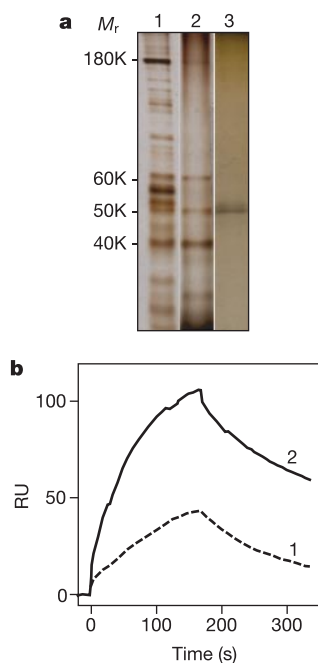


Figure 1 Affinity purification of the free apoA-I receptor. **a**, Total solubilized porcine liver plasma membrane proteins (lane 1) were purified either by apoA-I affinity chromatography (lane 2) or by micro-recovery on an apoA-I BI sensor chip (lane 3). Eluted proteins were resolved by SDS-PAGE and silver staining (lanes 1 and 2 contain 5 μ g of protein, lane 3 contains 0.1 μ g of protein). **b**, Biacore sensograms of the interaction of total solubilized porcine liver plasma membrane proteins injected either directly (curve 1), or after purification by apoA-I affinity chromatography (curve 2), with an apoA-I-bound sensor chip. Curves represent the resonance units as a function of time.

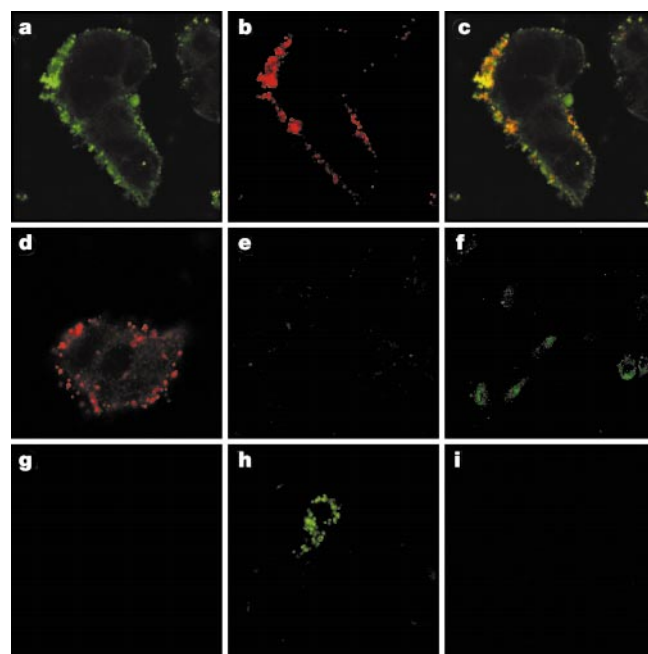


Figure 2 Immunofluorescence localization of the β - and α -chains of ATP synthase with apoA-I on the surface of hepatocytes. **a-c**, Colocalization of the β -subunit of ATP synthase (green, **a**) and apoA-I (red, **b**) on intact IHH cells is apparent in the merged image (yellow, **c**). **d**, Localization of the α -chain of ATP synthase by immunofluorescence using an antibody specific for the α -chain on intact IHH cells. **e**, **h**, As a control, an antibody against COX-I, another typical mitochondria protein, shows that COX-I is not present on intact (**e**) but is present on permeabilized (**h**) IHH cells. **f**, The β -subunit of ATP synthase is undetectable on intact CHO cells. **g**, **i**, Control experiments for antibody efficacy using isotypic purified mouse IgG (IgG2a) in IHH (**g**) and CHO (**i**) cells.

media (Methods). Notably, no synthesis of [32 P]ATP from ADP plus 32 P could be detected for up to 10 min either without or with free apoA-I (Supplementary Fig. 2a and b). By contrast, a hydrolysis activity was measured within 10 min (the generation of [α - 32 P]ADP from [α - 32 P]ATP), which was increased markedly (up to 79%) in the presence of apoA-I (Supplementary Fig. 2c and d). The purified F_1 protein, a natural inhibitor of mitochondrial F_1 -ATPase that interacts with the β -subunit to inhibit the ATP hydrolysis activity^{17,18}, could reduce both the basal (48% decrease as compared with the control) and the free apoA-I-stimulated hydrolysis activity (Supplementary Fig. 2e and f). Because ATP hydrolase activity is a main feature of a whole F_1 -ATPase complex⁵, our data indicate that on the plasma membrane of hepatocytes there is a complete F_1 -ATPase that functions as an ATP hydrolase and can be stimulated by free apoA-I.

We previously suggested that the ability of HDL or apoA-I to bind to high-affinity sites on hepatocytes might stimulate the internalization of HDL through low-affinity sites¹⁹. Indeed, whereas HDL₂ enriched in triglyceride (TG-HDL₂) binds to only low-affinity binding sites, remnant HDL₂ (a particle generated after the action of hepatic lipase on TG-HDL₂) binds to both low- and high-affinity binding sites and is internalized faster and to a greater extent than is wild-type TG-HDL₂ (ref. 19). Here we found that free apoA-I (0.1–20 μ g ml⁻¹) stimulated the internalization of 125 I-labelled TG-HDL₂ by HepG2 or IHH cells, with maximum internalization occurring between 5 and 15 min of incubation (data not shown).

Notably, 100 nM ADP (the only compound produced by ATP synthase at the hepatocyte cell surface) stimulated the internalization of TG-HDL₂ by 30% (Fig. 4a, column 2), similar to the amount stimulated by free apoA-I (Fig. 4a, column 1). The effect of both effectors was not additive (Fig. 4a, column 3), suggesting that they affect a similar endocytotic pathway. By contrast, 100 nM ATP stimulated only a small increase in TG-HDL₂ internalization (Fig. 4a, column 1). In addition, apyrase (E.C. 3.6.1.5), which hydrolyzes both ATP and ADP, completely abolished both the basal and the apoA-I-stimulated endocytosis of TG-HDL₂ (Fig. 4a, column 5). 2MeS-ADP, a non-hydrolysable analogue of ADP, showed a stimulatory effect at 10–100 nM, followed by an inhibition of endocytosis at higher concentrations (1–10 μ M). By contrast, ATP- γ S, a non-hydrolysable analogue of ATP, had a weak

effect (Fig. 4c). Together, these data indicate that there may be a specific ADP-dependent pathway for HDL endocytosis. Addition of epidermal growth factor (EGF), which induces endocytosis of the EGF receptor (EGFR), did not stimulate HDL internalization (Fig. 4a, column 6), indicating that HDL processing is not dependent on a nonspecific, general activation of endocytosis.

Endocytosis was not restricted to the protein moiety of HDL because free apoA-I also stimulated the internalization of [3 H]-cholesteryl-ether-labelled TG-HDL₂ (Fig. 4a, Chol). The ratio of [3 H]cholesteryl-ether to 125 I-labelled protein was also found to be the same in the original TG-HDL₂ solution and in the material recovered from inside the cells after a dissociation step, confirming that the whole HDL particle was involved in the endocytosis process.

HDL endocytosis could occur in hepatocytes through two different pathways: the first might be dependent on scavenger receptor class B type I (SR-BI), a widely described HDL receptor²⁰ that triggers internalization of the *holo*-HDL particle, followed by selective transcytosis of lipoprotein cholesterol; the second might be independent of SR-BI and involve the uptake and degradation of the *holo*-HDL particle by unknown receptors²¹. To evaluate the possible involvement of SR-BI in the stimulated HDL endocytosis, we carried out experiments using an IgG against SR-BI that can inhibit the uptake of cholesteryl ester in CHO SR-BI-transfected cells⁴ and in human adrenal NCI-H295R cells (V. Clavey, personal communication). This IgG had no effect on cholesteryl ester uptake

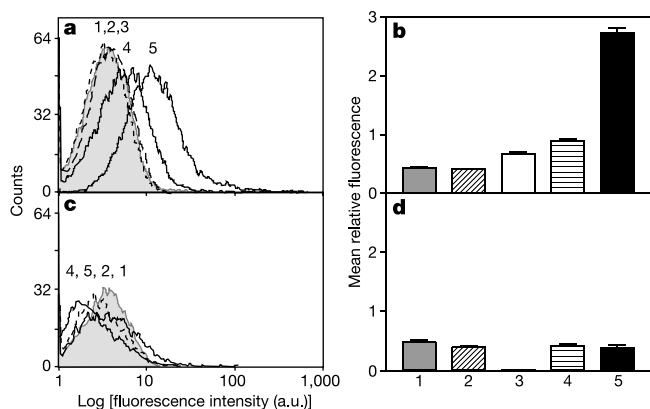


Figure 3 Detection of the β -chain of ATP synthase at the cell surface by flow cytometry. HepG2 and CHO cells were analysed by fluorescence-assisted flow cytometry. **a**, Analysis of HepG2 cells. Unbroken lines represent cells incubated with an antibody against the β -subunit of ATP synthase in the absence (curve 5) or presence of 100 nM apoA-I (curve 4) or 250 μ g ml⁻¹ of F_1 -ATPase (curve 3); broken lines represent cells incubated with an isotopic control mouse IgG2a (curve 1) or an antibody against COX-I (curve 2). **b**, Histogram showing the mean relative fluorescence of curves 1–5 in **a**. **c**, Analysis of CHO cells, as in described in **a** (curve 3 not done). **d**, Histogram showing the mean relative fluorescence of curves 1, 2, 4, 5 in **c**.

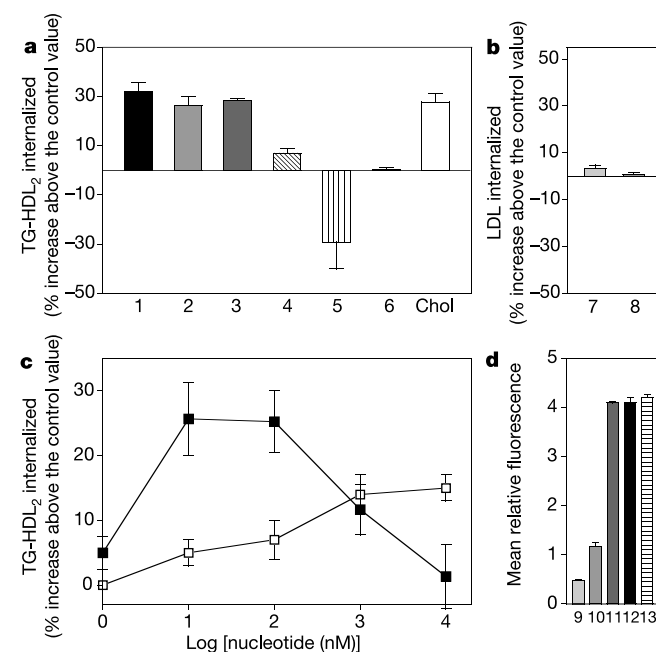


Figure 4 Effect of different nucleotides on internalization of TG-HDL₂ by hepatocytes. **a–c**, Cells were incubated for 10 min at 37 °C with 75 μ g ml⁻¹ 125 I-labelled TG-HDL₂ (**a**, **c**), [3 H]cholesteryl-ester-labelled TG-HDL₂ (**a**, Chol) or 125 I-labelled LDL (**b**). In some experiments, 10 μ g ml⁻¹ free apoA-I (columns 1 and 7), 100 nM ADP (columns 2 and 8), free apoA-I and ADP at the same concentrations as above (column 3), 100 nM ATP (column 4), 0.2 units ml⁻¹ apyrase (column 5), 100 nM EGF (column 6) or increasing concentrations of 2MeS-ADP (filled squares, **c**) and ATP- γ S (open squares, **c**) was added to the incubation medium. Data are expressed as a percentage above or below the control value (set as 0). **d**, Cells were pre-incubated for 10 min in serum-free medium with (column 10) or without (column 11) 20 nM EGF, or with 10 μ g ml⁻¹ free apoA-I (column 12) or 100 nM ADP (column 13). The amount of EGFR was measured by flow cytometry using antibodies against EGFR. An isotopic IgG was also used as a control for the efficacy of the EGFR antibody (column 9). Results are expressed as the mean relative cell fluorescence; receptor internalization is indicated by a drop in fluorescence.

in hepatocytes, thus cell-surface SR-BI is not involved in the early events of HDL endocytosis (data not shown).

A principal feature of the observed endocytosis is its specificity towards HDL. Indeed, endocytosis of ^{125}I -labelled LDL was not stimulated by either apoA-I or ADP (Fig. 4b, columns 7 and 8, respectively). Using flow cytometry, we also measured internalization of the EGFR, as representative of a tyrosine kinase receptor on HepG2 (Fig. 4d) or IHH cells (data not shown). Whereas EGF strongly reduced the number of EGFRs at the cell surface (Fig. 4d, column 10) as compared with control cells (Fig. 4d, column 11), free apoA-I (Fig. 4d, column 12) or ADP (Fig. 4d, column 13) had no effect on the number of EGFRs on HepG2 cells. Increasing concentrations of IF $_1$ protein (a natural inhibitor of the ATP hydrolysis activity of mitochondrial ATP synthase²²) strongly inhibited both the basal (Fig. 5a) and the free apoA-I-stimulated (Fig. 5b) TG-HDL $_2$ endocytosis, strengthening the idea that ADP generation has a chief role in basal endocytosis or in endocytosis stimulated by high-affinity receptors.

The ATP-binding cassette A-1 (ABCA-1) which binds free apoA-I (ref. 23) and is present on HepG2 and IHH cells, might be a partner in this process. Although no binding parameters for ABCA-1 have been measured in this particular cell line²⁴, we extrapolated from binding data in other cell types the binding parameter of free apoA-I on ABCA-1 (refs 25, 26). It would represent about 4–10% of the total free apoA-I binding (25 ng per mg of cell protein) observed under our conditions on HepG2 cells and IHH, and thus would be almost undetectable. In addition, we observed the stimulation (2–2.5-fold) of cholesterol efflux induced by free apoA-I (a typical feature of ABCA-1; ref. 24), but this was not influenced by IF $_1$ or ADP (over the range 1 nM to 10 μM), which allowed us to conclude that our observations did not involve ABCA-1.

To estimate the physiological relevance of our data, we carried out *in situ* internalization experiments using perfused rat liver (ref. 27 and Supplementary Table I). Notably, IF $_1$ protein induced a rapid (45 min) and marked decrease (up to 45%) of TG-HDL $_2$ internalization by the liver, indicating that, at least in rodent, the ectopic ATP synthase seems to be implicated in hepatic HDL endocytosis.

The role that we have proposed for cell-surface ATP synthase in HDL catabolism raises possibilities for the control of cholesterolemia—a fundamental issue in cardiovascular disease research. Our

findings provide a sequence of events whereby the membrane-bound F $_1$ -ATPase elicits a cellular response by generating extra-cellular ADP with the probable involvement of specific downstream receptors. With regard to HDL, these mechanisms seem to be part of a regulation pathway, because the high-affinity binding of HDL to the β -chain strongly stimulates ADP generation, which in turn promotes HDL endocytosis. But how the cell directs these proteins towards the cell surface and how their cell-surface expression (which seems to be restricted to particular types of cell) is regulated remain unknown and require further investigation. □

Methods

Receptor purification using surface plasmon resonance

Surface plasmon resonance measurements and recovery were done at 20 °C using a Biacore 3000 (Biacore AB) equipped with a research-grade B1 sensor chip (Biacore AB). ApoA-I (50 fmol mm $^{-2}$) was immobilized on three flow cells using traditional amine-coupling chemistry²⁸. The fourth control flow cell lacked immobilized apoA-I. We diluted solubilized porcine liver plasma membrane proteins (in 125 mM Tris maleate, 1 mM CaCl $_2$, 150 mM NaCl and 8 mM CHAPS, pH 7.4) eightfold in HBS running buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA and 0.005% polysorbate, pH 7.4), and injected 100 μg of proteins at a flow rate of 20 $\mu\text{l min}^{-1}$. The APROG microrecovery procedure (Biacore AB) was used to recover captured proteins in 7 μl of elution buffer (10 mM triethylamine and 6 M urea, pH 11). We resolved the eluates by SDS-PAGE and silver staining. To measure the binding activity of porcine liver plasma membrane proteins, the proteins eluted from apoA-I affinity chromatography were diluted threefold in running buffer, and 1.5 μg of proteins was injected into either the apoA-I cell or the control flow cell.

Affinity chromatography and sequence analysis

Apolipoprotein AI, coupled to an Affi-gel 15 support (Bio-Rad), was used to purify apoA-I-binding proteins. Solubilized porcine liver plasma membrane extracts were diluted to a final concentration of 1 mM in CHAPS buffer and applied to the apoA-I affinity column for 1 h at 4 °C. After five column washes with 10 ml 0.1 M sodium acetate buffer, pH 6.5, bound proteins were eluted in 10 mM triethylamine and 6 M urea, pH 11. We resolved the eluates by SDS-PAGE. An amidoblack-stained band of 50K was cut out and digested with endoprotease lysine-C. The resulting peptides were separated by HPLC on a C18 column with a 2–70% gradient of acetonitrile in 0.1% trifluoroacetic acid and sequenced at the Institut Pasteur.

Internalization assays

Internalization assays were done as described¹⁹. Results are expressed as a percentage of the internalization measured with ^{125}I -TG-HDL $_2$ alone (corresponding to a value of 400 ng of TG-HDL $_2$ per mg of cell protein).

Flow cytometry

The primary antibodies were from Molecular Probes as follows: clone 7E3-F2 against the β -chain of ATP synthase, 7H10-BD4 against the α -chain of ATP synthase, and 1D6-E1-A8 against subunit I of cytochrome oxidase. Antibodies (clone LA1) against EGFR were from Upstate Biotechnology.

To analyse cell-surface EGFR, HepG2 cells were preincubated in medium with or without 20 nM EGF for 10 min at 37 °C. For flow cytometry analysis, HepG2 and CHO cells were detached and fixed in 3% paraformaldehyde. We incubated cells at 20 °C for 1 h in PBS, pH 7.4, containing 1% bovine serum albumin (BSA) plus primary monoclonal antibodies. The cells were then washed in PBS plus 1% BSA and incubated at 20 °C for 30 min with a goat antibody against mouse IgG conjugated to fluorescein isothiocyanate, before analysis on a Coulter XL 4C flow cytometer (Beckman-Coulter).

Immunofluorescence and confocal microscopy

Glass coverslips coated with cells were washed with PBS, pH 7.4, fixed for 15 min in 3% paraformaldehyde and saturated for 30 min with 0.2% gelatin (staining buffer). A control slide was permeabilized for 2 min in 0.2% Triton X-100. Cells were then incubated for 1 h with the primary antibody diluted to 5 $\mu\text{g ml}^{-1}$ in PBS. Immunostaining was carried out for 1 h in the dark with antibody against mouse IgG2a conjugated to Alexa 488 diluted to 5 $\mu\text{g ml}^{-1}$ in staining buffer.

For confocal microscopy, we incubated the cells for 2 h with 100 $\mu\text{g ml}^{-1}$ apoA-I, and then washed them twice in PBS before fixation. A rabbit polyclonal antiserum against apoA-I (10 $\mu\text{g ml}^{-1}$) was incubated with the primary antibodies as described above. Immunostaining was done with an antibody against mouse IgG2a conjugated to Alexa 488 (5 $\mu\text{g ml}^{-1}$) and an antibody against rabbit IgG conjugated to rhodamine (5 $\mu\text{g ml}^{-1}$). The coverslips were examined with a Zeiss Axioskop microscope or with a confocal microscope (LSM510, Zeiss) at a magnification of $\times 630$.

Measurement of cell-surface ADP and ATP

Confluent HepG2 cells in six-well plates were washed in DMEM medium, and then incubated at 37 °C for 10 min in DMEM, pH 6.6, with 0.1 μCi of [α - ^{32}P]ATP for the ADP generation assay, or with 0.1 μCi ^{32}P and ADP (100 nM final concentration) for the ATP generation assay. Depending on the experiment, IF $_1$ (100 nM final) or apoA-I (10 $\mu\text{g ml}^{-1}$ final) was added to the reaction mixture. We analysed the supernatants by two systems: first, by HPLC coupled to a radioactivity detector on a Whatman Partisphere 5 SAX column as described²⁹, with calibration by radiolabelled nucleotides; second, by thin-layer

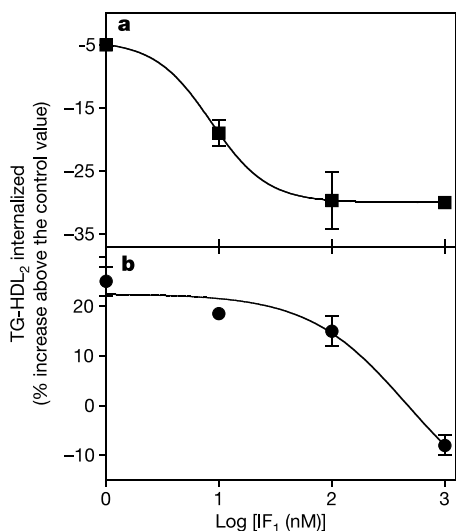


Figure 5 Effect of IF $_1$ on TG-HDL $_2$ internalization by hepatocytes. HepG2 cells were incubated for 10 min at 37 °C in DMEM medium, pH 6.6, with 75 $\mu\text{g ml}^{-1}$ of ^{125}I -labelled TG-HDL $_2$ and in the presence of increasing concentrations of IF $_1$ without (a) or with (b) free apoA-I (10 $\mu\text{g ml}^{-1}$). Data are expressed as a percentage above or below the control value (set as 0).

chromatography in 2.4% NaCl:NH₄OH:H₂O:MeOH (12.5:15:27.5:50 v/v) with quantification of radioactive spots by liquid scintillation.

Received 30 July; accepted 7 October 2002; doi:10.1038/nature01250.

1. Sviridov, D. & Nestel, P. Dynamics of reverse cholesterol transport: protection against atherosclerosis. *Atherosclerosis* **161**, 245–254 (2002).
2. Fidge, N. H. High density lipoprotein receptors, binding proteins, and ligands. *J. Lipid Res.* **40**, 187–201 (1999).
3. Barbaras, R., Collet, X., Chap, H. & Perret, B. Specific binding of free apolipoprotein A-I to a high-affinity binding site on HepG2 cells: characterization of two high-density lipoprotein sites. *Biochemistry* **33**, 2335–2340 (1994).
4. Martinez, L. O. *et al.* Characterization of two high-density lipoprotein binding sites on porcine hepatocyte plasma membranes: contribution of scavenger receptor class B type I (SR-BI) to the low-affinity component. *Biochemistry* **39**, 1076–1082 (2000).
5. Boyer, P. D. The ATP synthase: a splendid molecular machine. *Annu. Rev. Biochem.* **66**, 717–749 (1997).
6. Stock, D., Leslie, A. G. & Walker, J. E. Molecular architecture of the rotary motor in ATP synthase. *Science* **286**, 1700–1705 (1999).
7. Cabezon, E., Runswick, M. J., Leslie, A. G. & Walker, J. E. The structure of bovine IF₁, the regulatory subunit of mitochondrial F₁-ATPase. *EMBO J.* **20**, 6990–6996 (2001).
8. Das, B., Mondragon, M. O., Sadeghian, M., Hatcher, V. B. & Norin, A. J. A novel ligand in lymphocyte-mediated cytotoxicity: expression of the beta subunit of H⁺ transporting ATP synthase on the surface of tumour cell lines. *J. Exp. Med.* **180**, 273–281 (1994).
9. Moser, T. L. *et al.* Angiotensin binds ATP synthase on the surface of human endothelial cells. *Proc. Natl Acad. Sci. USA* **96**, 2811–2816 (1999).
10. Chang, S. Y., Park, S. G., Kim, S. & Kang, C. Y. Interaction of the C-terminal domain of p43 and the α subunit of ATP synthase: Its functional implication in endothelial cell proliferation. *J. Biol. Chem.* **277**, 8388–8394 (2002).
11. Schippers, I. J. *et al.* Immortalized human hepatocytes as a tool for the study of hepatocytic (de-)differentiation. *Cell Biol. Toxicol.* **13**, 375–386 (1997).
12. Williams, N., Amzel, L. M. & Pedersen, P. L. Proton ATPase of rat liver mitochondria: a rapid procedure for purification of a stable, reconstitutively active F₁ preparation using a modified chloroform method. *Anal. Biochem.* **140**, 581–588 (1984).
13. Moser, T. L. *et al.* Endothelial cell surface F₁-F₀ ATP synthase is active in ATP synthesis and is inhibited by angiotensin. *Proc. Natl Acad. Sci. USA* **98**, 6656–6661 (2001).
14. Beisiegel, U. *et al.* Apolipoprotein E-binding proteins isolated from dog and human liver. *Arteriosclerosis* **8**, 288–297 (1988).
15. Mahley, R. W., Hui, D. Y., Innerarity, T. L. & Beisiegel, U. Chylomicron remnant metabolism. Role of hepatic lipoprotein receptors in mediating uptake. *Arteriosclerosis* **9**, 114–118 (1989).
16. Ferrini, J. B., Pichard, L., Domergue, J. & Maurel, P. Long-term primary cultures of adult human hepatocytes. *Chem. Biol. Interact.* **107**, 31–45 (1997).
17. Cabezon, E., Arechaga, I., Jonathan, P., Butler, G. & Walker, J. E. Dimerization of bovine F₁-ATPase by binding the inhibitor protein, IF₁. *J. Biol. Chem.* **275**, 28353–28355 (2000).
18. Cabezon, E., Butler, P. J., Runswick, M. J. & Walker, J. E. Modulation of the oligomerization state of the bovine F₁-ATPase inhibitor protein, IF₁, by pH. *J. Biol. Chem.* **275**, 25460–25464 (2000).
19. Guendouzi, K., Collet, X., Perret, B., Chap, H. & Barbaras, R. Remnant high density lipoprotein2 particles produced by hepatic lipase display high-affinity binding and increased endocytosis into a human hepatoma cell line (HEPG2). *Biochemistry* **37**, 14974–14980 (1998).
20. Acton, S. *et al.* Identification of scavenger receptor SR-BI as a high density lipoprotein receptor. *Science* **271**, 518–520 (1996).
21. Silver, D. L., Nan, W., Xiao, X. & Tall, A. R. HDL particle uptake mediated by SR-BI results in selective sorting of HDL cholesterol from protein and polarized cholesterol secretion. *J. Biol. Chem.* **276**, 25287–25293 (2001).
22. Walker, J. E. The regulation of catalysis in ATP synthase. *Curr. Opin. Struct. Biol.* **4**, 912–918 (1994).
23. Wang, N., Silver, D. L., Thiele, C. & Tall, A. R. ATP-binding cassette transporter A1 (ABCA1) functions as a cholesterol efflux regulatory protein. *J. Biol. Chem.* **276**, 23742–23747 (2001).
24. Bortnick, A. E. *et al.* The correlation of ABCA1 mRNA levels with cholesterol efflux from various cell lines. *J. Biol. Chem.* **275**, 28634–28640 (2000).
25. Burgess, J. W., Kiss, R. S., Zheng, H., Zachariah, S. & Marcel, Y. L. Trypsin-sensitive and lipid-containing sites of the macrophage extracellular matrix bind apolipoprotein A-I and participate in ABCA1-dependent cholesterol efflux. *J. Biol. Chem.* **277**, 31318–31326 (2002).
26. Fitzgerald, M. L. *et al.* Naturally occurring mutations in ABCA1's largest extracellular loops can disrupt its direct interaction with apolipoprotein A-I. *J. Biol. Chem.* **277**, 33178–33187 (2002).
27. Barrans, A. *et al.* Hepatic lipase induces the formation of pre- β 1 high density lipoprotein (HDL) from triacylglycerol-rich HDL₂. *J. Biol. Chem.* **269**, 11572–11577 (1994).
28. Johnsson, B., Lofas, S. & Lindquist, G. Immobilization of proteins to a carboxymethyl-dextran-modified gold surface for biospecific interaction analysis in surface plasmon resonance sensors. *Anal. Biochem.* **198**, 268–277 (1991).
29. Sultan, C. *et al.* The novel inositol lipid phosphatidylinositol 3,4-bisphosphate is produced by human blood platelets upon thrombin stimulation. *Biochem. J.* **269**, 831–834 (1990).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank G. Larrieu, for technical help, and P. Maurel INSERMU128, for the primary human hepatocytes.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.B. (e-mail: Ronald.Barbaras@toulouse.inserm.fr).

Chloroplast to nucleus communication triggered by accumulation of Mg-protoporphyrinIX

Åsa Strand*†, Tadao Asami‡, Jose Alonso*, Joseph R. Ecker* & Joanne Chory*†

* Plant Biology Laboratory; and † the Howard Hughes Medical Institute; and The Salk Institute, La Jolla, California 92037, USA

‡ Plant Functions Laboratory, Riken, Wako, Saitama 351-0198, Japan

Plant cells coordinately regulate the expression of nuclear and plastid genes that encode components of the photosynthetic apparatus. Nuclear genes that regulate chloroplast development and chloroplast gene expression provide part of this coordinate control. There is evidence that information also flows in the opposite direction, from chloroplasts to the nucleus^{1,2}. Until now, at least three different signalling pathways have been identified that originate in the plastid and control nuclear gene expression^{3,4} but the molecular nature of these signals has remained unknown. Here we show that the tetrapyrrole intermediate Mg-protoporphyrin (Mg-ProtoIX) acts as a signalling molecule in one of the signalling pathways between the chloroplast and nucleus. Accumulation of Mg-ProtoIX is both necessary and sufficient to regulate the expression of many nuclear genes encoding chloroplastic proteins associated with photosynthesis.

Communication between plastids and the nucleus is necessary for the initiation of chloroplast development in the light, and also for the ability of the plant cell to respond correctly to fluctuations in the environment. Evidence that nuclear genes are regulated by signals originating in the plastid came from studies with mutants of carotenoid biosynthesis. These mutants bleach when exposed to high irradiances of light, and show decreased expression of nuclear photosynthetic genes⁵. This photobleaching response was later replicated by treating wild-type plants with norflurazon, a non-competitive inhibitor of carotenoid biosynthesis⁶. Norflurazon-treated plants suffer from photooxidation of the thylakoid membranes and the resulting inhibition of chloroplast development and function leads to decreased transcription of nuclear encoded-photosynthetic genes^{7,8}. Using the effect of norflurazon on nuclear gene expression, we have previously undertaken a genetic approach to identify components of the plastid to nucleus signalling pathway(s)⁹. Five non-allelic loci were identified as genome uncoupled mutants (*gun1–5*); these mutants express nuclear-encoded photosynthetic genes in the absence of proper chloroplast development. Three of these loci encode enzymes in the tetrapyrrole pathway; *GUN2* encodes haem oxygenase (allelic to *HY1*), *GUN3* phytylcholine synthase (allelic to *HY2*), and *GUN5* the H-subunit of Mg-chelatase (Fig. 2)³. Here we use the *gun1*, *gun2* and *gun5* mutants, as well as new mutants, to unravel the source of the plastid signal regulating nuclear-encoded genes.

To further characterize the extent of nuclear genes whose expression is regulated by the plastid signal, we used the Affymetrix *Arabidopsis* oligoarray, containing about 8,200 genes, and showed that, in addition to the previously described photosynthetic genes (*LHCB* and *RBCS*), 322 genes change their expression more than threefold (182 repressed and 140 induced), in wild-type seedlings grown on 5 μ M norflurazon (Supplementary Information). Genetic analysis of the different *gun* mutants suggests there are two separate *gun* pathways, defined by mutations in *GUN1* and *GUN2–5* (ref. 3), so we compared the expression profiles of three *gun* mutants, *gun1*, *gun2* and *gun5*. Of the 322 genes whose expression levels were

affected by norflurazon treatment in wild type, 152 showed more than a threefold difference in one or more of the *gun* mutants relative to wild type. Cluster analysis of those 152 genes showed that the expression profiles of the *gun2* and *gun5* mutants clustered together (Fig. 1a), whereas *gun1* showed a different gene expression profile. These data strongly support the genetic analysis that shows *GUN1* is involved in a signalling pathway separate from the one defined by *GUN2–5* (Fig. 1a). Of the 182 genes repressed by norflurazon treatment in wild type, about 70 genes were not repressed in *gun2* and *gun5*. All the proteins encoded by the genes with known function are predicted to be localized to the chloroplast. These genes were further categorized as encoding proteins involved in the light-harvesting and electron-transport reactions of photosynthesis, enzymes in metabolic pathways such as the Calvin cycle and tetrapyrrole biosynthesis, and proteins involved in translation of chloroplast-encoded genes (Fig. 1b). Thus, the *GUN2–5* pathway affects a large number of nuclear genes encoding chloroplastic proteins directly associated with photosynthetic reactions.

The similar expression profiles for the *gun2* and *gun5* mutants, and the fact that both mutants have lesions in enzymes involved in tetrapyrrole biosynthesis strongly suggested that perturbations of the tetrapyrrole pathway were associated with the plastid signal. In the *gun5* mutant, Mg-chelatase activity is reduced and as a result *gun5* seedlings contain approximately 75% of the wild-type amount of chlorophyll (Table 1)³. The *gun2* seedlings had an even paler phenotype (Table 1), partly owing to feedback inhibition of early steps in the tetrapyrrole pathway caused by haem accumulation¹⁰. To examine whether norflurazon-treated seedlings maintain a capacity for tetrapyrrole biosynthesis, despite the reduced amount of chlorophyll in these seedlings (Table 1), aminolevulinic acid (ALA), an early precursor for tetrapyrroles (Fig. 2), was fed to green and norflurazon-treated seedlings and the amounts of protoporphyrinIX (ProtoIX) and Mg-ProtoIX were analysed using fluorescence. The norflurazon-grown seedlings accumulated tenfold higher amounts of Mg-ProtoIX following ALA feeding compared to green seedlings (data not shown) suggesting that norflurazon-grown seedlings maintain flux through the tetrapyrrole pathway but do not have the capacity to make chlorophyll. The accumulation of Mg-ProtoIX in norflurazon-grown seedlings can be explained by the fact that the latter steps in chlorophyll biosynthesis are localized

to the thylakoid membranes and that those structures are destroyed by the effects of norflurazon treatment¹¹. We investigated whether Mg-ProtoIX also accumulates in the norflurazon-grown seedlings under normal growth conditions, without ALA feeding by high-performance liquid chromatography (HPLC) analysis. The norflurazon-grown wild-type seedlings did indeed show a several-fold increase of Mg-ProtoIX compared to the green seedlings (Table 1). The *gun2* and *gun5* mutants accumulated significantly less than wild type (Table 1) owing to reduced flux through the tetrapyrrole pathway in these mutants. These data suggest a correlation between accumulation of Mg-ProtoIX and reduced expression of photosynthetic genes. The photoreactive properties of tetrapyrroles made us investigate whether the difference in gene regulation in the *gun* mutants was a consequence of reduced production of reactive oxygen species. Norflurazon-grown seedlings from wild type, *gun2* and *gun5* were stained for superoxide using nitroblue tetrazolium¹² to assess whether the reduced accumulation of Mg-ProtoIX in the *gun* mutants results in generation of less reactive oxygen species. The *gun* mutants stain at least as strongly as wild type, suggesting there is no difference in the amount of reactive oxygen species produced in wild type and the *gun* mutants (data not shown).

To test the idea further that accumulation of Mg-ProtoIX is the trigger for repressing nuclear gene expression, inhibitors were used to manipulate the tetrapyrrole pathway. Specifically, we wished to examine the effects of Mg-ProtoIX on photosynthetic gene expression in *Arabidopsis*, and whether the reduced accumulation of Mg-ProtoIX in *gun2* and *gun5* was responsible for the *gun* phenotype. Dipyridyl (DP), which is an Fe-chelator known to increase the levels Mg-ProtoIX and its methyl esters¹³, was used to force *gun2* and *gun5* mutants to accumulate higher amounts of Mg-ProtoIX. When these norflurazon grown *gun* mutants were treated with DP for 8 h the *gun* phenotype was lost (Fig. 2a), strongly suggesting that Mg-ProtoIX accumulation is required for chloroplast-to-nucleus signalling. To demonstrate that the effect was not caused by toxicity of DP we also used S23142, a potent inhibitor of protoporphyrinogen oxidase¹⁴. In the presence of both inhibitors, Mg-ProtoIX does not accumulate because S23142 blocks the enzymatic reaction upstream of Mg-chelatase. Under these conditions, both *gun2* and *gun5* mutants retained the *gun* phenotype (Fig. 2a).

If the reduced accumulation of Mg-ProtoIX is the reason photosynthetic genes are misregulated during norflurazon treatment in *gun2* and *gun5*, other mutants in the tetrapyrrole pathway should also exhibit *gun* phenotypes. We identified knockout mutants for porphobilinogen deaminase (PBD) and the D-subunit of Mg-chelatase (see Methods). We also obtained a mutant for coproporphyrinogen oxidase, *lin2* (ref. 26), and constructed plants co-suppressed for the H-subunit of Mg-chelatase. Owing to reduced flux through the pathway, the mutants with different lesions in the tetrapyrrole biosynthesis pathway should not accumulate Mg-ProtoIX up to the critical threshold level, as would the wild type (Fig. 2).

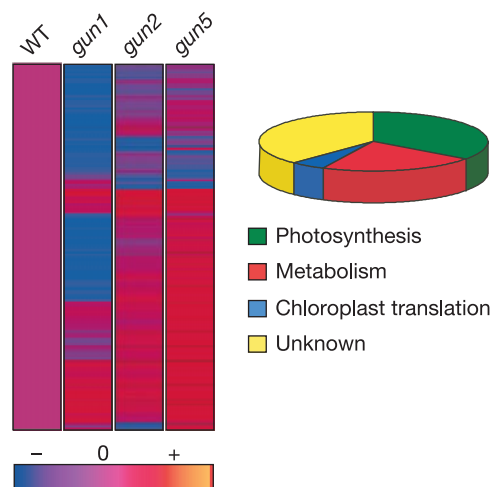


Figure 1 Micorarray analysis demonstrates that a large number of nuclear encoded genes are regulated by the plastid signal. **a**, Cluster analysis of norflurazon-regulated genes induced or repressed by at least three fold in *gun1*, *gun2* and *gun5* compared to wild type. Red colour represents an increase in expression and blue colour a decrease relative to wild type. **b**, Classification of the genes that were not repressed in *gun2* and *gun5* compared to wild type when grown on norflurazon.

Table 1 Chlorophyll, Mg-ProtoIX and Mg-ProtoIX ME content in various genotypes

	Wild type Green	<i>gun2</i> Green	<i>gun5</i> Green	Wild type NF	<i>gun2</i> NF	<i>gun5</i> NF
Chl (nmol g ⁻¹ FW)	837	411	667	ND	ND	ND
Mg-ProtoIX (pmol g ⁻¹ FW)	414	360	16	6,380	1,730	344
Mg-ProtoIX ME (pmol g ⁻¹ FW)	67	149	23	24	13	ND

The increased accumulation of Mg-ProtoIX in the norflurazon-grown wild-type seedlings compared to green seedlings was replicated in five different experiments and were consistently higher (4–15-fold) in norflurazon-treated wild type. Extracts containing detectable quantities of tetrapyrroles are difficult to obtain so the total amount of Mg-ProtoIX and Mg-ProtoIX ME varied between experiments; we therefore show data from a single experiment, in which all samples were harvested and quantified on the same day. CoproporphyrinogenIII, protoporphyrinogenIX, protoporphyrinIX and protochlorophyllide could be determined by the HPLC system used but none of those intermediates could be detected in norflurazon-treated plants (data not shown). The reduced accumulation of Mg-ProtoIX in *gun2* and *gun5* compared to wild type was repeated twice with similar results. ND, not detected; NF, norflurazon; FW, fresh weight; ME, methyl ester.

If our model is correct, *LHCB* expression would not be repressed in these backgrounds. All of these mutants accumulated *LHCB* messenger RNA when grown on norflurazon (Fig. 2b). Thus, we have identified a number of new *gun* mutants.

To directly demonstrate a role for Mg-ProtoIX as a regulator of nuclear genes, protoplasts were prepared from *Arabidopsis* leaves and incubated with Mg-ProtoIX for 5 h. *LHCB* expression was significantly repressed after treatment with Mg-ProtoIX (Fig. 2c), thereby providing strong evidence that accumulation of Mg-ProtoIX is a negative regulator of *LHCB* expression. Feeding of porphobilinogen, haem and protoporphyrinIX to the protoplasts showed no effect on *LHCB* transcript levels, demonstrating that feeding of tetrapyrroles is not toxic and that the effect on *LHCB* expression is specific for Mg-ProtoIX (Fig. 2c). Taken together, our data provide compelling evidence that accumulation of Mg-ProtoIX is both necessary and sufficient to repress the expression of a large number of nuclear genes during conditions where chlorophyll synthesis is inhibited.

To define the target of the Mg-ProtoIX signal we examined characterized elements in the promoter of one gene known to be affected by norflurazon, *LHCB1*. Two of the best-defined binding sites involved in light-regulated transcription of *LHCB* genes are the GT-1 (G3M) and the G-box (CUF1) elements¹⁵. To investigate if either element responds to the Mg-ProtoIX-mediated signal we used three different *LHCB1::Luciferase* reporter constructs that are

truncated or contain mutations of the two key elements, GT-1 and CUF1 (Fig. 3)¹⁶. These three different *LHCB1::Luciferase* constructs were transformed into wild-type plants¹⁶ and subsequently crossed into the *gun5* mutant. Luciferase activity was measured after growth of the transgenic lines in either wild-type or *gun5* backgrounds on norflurazon. In the CUF1M lines there was no difference in luciferase activity between plants with either wild-type or *gun5* backgrounds (Fig. 3c). This is in contrast to the G3M and -142 lines where the *gun* phenotype was clearly observed (Fig. 3a and b). This result demonstrates that the G-box motif defined by CUF1 responds to the Mg-ProtoIX-mediated signal (Fig. 3c). A true CUF1 (CACGTA) element is present in the promoter region in 18 of the 70 genes regulated by the plastid signal (Fig. 1), additionally 24 genes have the closely related CACGTG (data not shown).

On the basis of our results, we propose a model for the Mg-ProtoIX mediated plastid-to-nucleus communication (Fig. 4). Under normal chloroplast development transcription of the *LHCB1* gene is partly regulated through the CUF1 element, either by an activator stimulating expression or by inhibition of a repressor (Fig. 4a and c). Under conditions where the latter steps in chlorophyll biosynthesis are inhibited, Mg-ProtoIX accumulates in the chloroplasts and eventually diffuses or is transported to the cytosol. In the cytosol, Mg-ProtoIX is bound by a regulatory protein, possibly a transcription factor, and modifies the activity and/or the translocation of this protein. As a consequence, expression of photosynthetic genes is inhibited (Fig. 4b and d). In *gun5* the activation of *LHCB1* is maintained or the binding of the repressor does not occur owing to reduced accumulation of Mg-ProtoIX. Future work investigating *trans*-acting factors will help distinguish between the two possible regulatory mechanisms triggered by Mg-ProtoIX accumulation.

In *Chlamydomonas*, Mg-ProtoIX and its methyl esters have been shown to substitute for light in the induction of the nuclear *HSP70* gene encoding a chloroplast-localized chaperone^{17,18}. Furthermore,

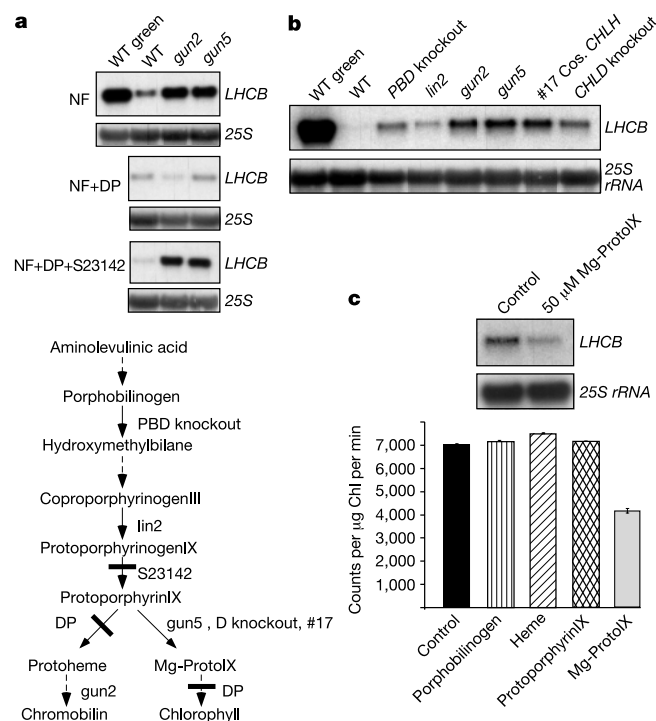


Figure 2 Northern blot analysis indicates a role for Mg-ProtoIX as a negative regulator of *LHCB* expression. **a**, *LHCB* expression in *Arabidopsis* seedlings grown in liquid culture after norflurazon, DP and S23142 treatment. **b**, *LHCB* expression in different mutants in the tetrapyrrole pathway. Lesions in porphobilinogen deaminase (PBD KO), coproporphyrinogen oxidase (*lin2*), haem oxygenase (*gun2*), the H-subunit of Mg-chelatase (*gun5* and #17 co-suppressed line) and the D-subunit of Mg-chelatase (D KO) all result in a *gun* phenotype when grown on 5 μ M norflurazon. **c**, Feeding of 50 μ M Mg-ProtoIX to protoplasts for 5 h in low light results in reduced expression levels of *LHCB*. This was demonstrated both by northern blot analysis and luciferase activity in protoplasts prepared from *Arabidopsis* plants containing a *LHCB1::Luciferase* reporter constructs. Feeding of 50 μ M porphobilinogen, haem and protoporphyrinIX to the protoplasts showed no effect on *LHCB* transcript levels. All the northern blot analyses presented were repeated with identical results. 25S rRNA is a loading control.

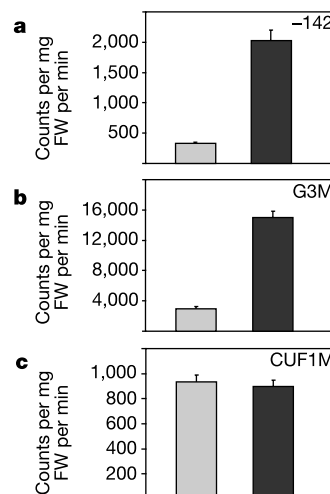
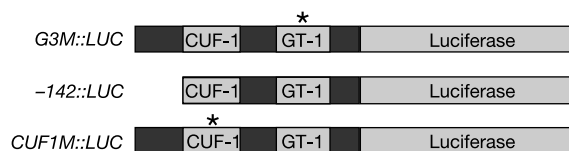


Figure 3 The Mg-ProtoIX-mediated signal is received by the CUF element of the *LHCB1* promoter. Luciferase assays carried out on wild-type and *gun5* lines that contain different *LHCB1::Luciferase* reporter constructs. The values presented are means from 12 different samples from each genotype collected after growth on 5 μ M norflurazon. Wild type is shown with shaded bars and *gun5* with filled bars. Asterisks show which element has been mutated in the two different constructs, G3M and CUF1M.

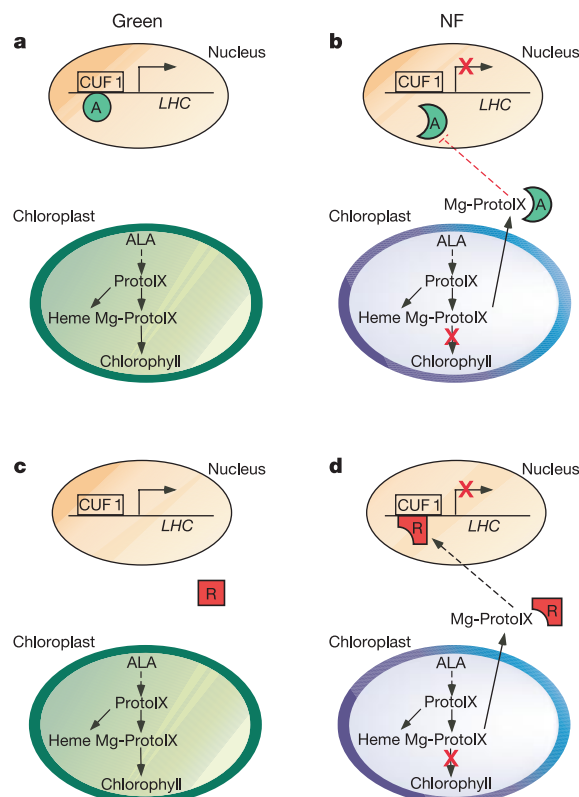


Figure 4 Models for the Mg-ProtoIX mediated plastid signal. Transcription of the *LHCBI* gene is partly regulated through the CUF1 element, either by an activator stimulating expression or by inhibition of a repressor (Fig. 4a and c). However, under conditions where the latter steps in chlorophyll biosynthesis are inhibited, Mg-ProtoIX accumulates in the chloroplasts and is thought to be exported to the cytosol where it mediates a signal to the nucleus to repress the expression of photosynthetic genes. Mg-ProtoIX accumulation either inhibits a transcription factor activating *LHCBI* expression or stimulates the binding of a negative regulator (Fig. 4b and d).

treatment of *Chlamydomonas* cells with the inhibitor DP results in decreased expression of *LHCB* genes¹⁹. There are also reports from experiments on higher plants suggesting that perturbations in the tetrapyrrole pathway might affect *LHCB* expression^{8,20}. However, a definitive role for tetrapyrroles in regulation of nuclear-encoded genes has so far been elusive^{1,2,21}. Our *gun* mutants with defects in defined steps in tetrapyrrole biosynthesis have enabled us to identify Mg-ProtoIX as a signalling molecule in one of the plastid-to-nucleus retrograde signalling pathways. In yeast it has been demonstrated that molecular oxygen and haeme synthesized in the mitochondria regulate transcription of nuclear genes that encode mitochondrial proteins²². Because plants contain both chloroplasts and mitochondria, it is sensible that the chloroplast communicates with the nucleus via a plastid-specific molecule such as Mg-ProtoIX. □

Methods

Plant material and growth conditions

For all experiments, *Arabidopsis* wild-type and mutant seedlings were grown for 6 days in continuous white light of 300 $\mu\text{mol m}^{-2} \text{s}^{-1}$ on 1 $\times$ MS (MS, Murashige Skoog medium), 2% sucrose plates with and without 5 μM norflurazon (Sandoz Pharmaceutical). For the inhibitor studies, seedlings were grown in liquid cultures (1 $\times$ MS, 2% sucrose and 5 μM norflurazon) for 6 days, 5 mM DP (Sigma) and 1 μM of the diphenylether herbicide S23142 (ref. 14) was added to the cultures and incubated for an additional 8 h. This time period is enough to induce the plastid signal in the *gun* mutants. However, 8 h incubation is not sufficient to remove the signal in wild type and to accumulate detectable levels of *LHCB* transcript. Protoplast isolation was performed as in ref. 23 and the protoplasts incubated in low light with 50 μM Mg(II) protoporphyrin IX disodium salt, Haemin (Ferriprotoporphyrin IX chloride), porphobilinogen monohydrate or protoporphyrin IX

(Frontier Scientific) for 5 h. The PBD knockout and CHLD knockout have transfer DNA insertions in the second and the fourth exons, respectively. Homozygous lines for either knockout mutant are unable to synthesise chlorophyll and are consequently not able to establish growth beyond the seedling stage. Seeds were therefore germinated on MS-plates and after two days, homozygous lines were selected by their pale phenotype and transferred to norflurazon plates. The same treatment was performed for wild type. To create plants with reduced levels of ChlH protein by cosuppression, a complementary DNA for the H-subunit of Mg-chelatase was obtained by PCR with reverse transcription (RT-PCR) and cloned into the binary vector pCHF1 used to transform *Arabidopsis*.

RNA isolation and *Arabidopsis* oligoarrays

Arabidopsis seedlings were harvested in the light and RNA was prepared using the Plant RNeasy Kit (QIAGEN). Probe preparation, hybridization to Affymetrix *Arabidopsis* oligoarrays, washing, staining and scanning were carried out according to the manufacturer's suggestions (Affymetrix). The data analysis was performed using GeneSpring 4.2 (Silicon Genetics). The *Arabidopsis* annotation database from the laboratory of J. Schroeder was used²⁴. To ensure data quality only genes with an expression level over 300 (raw data) were used for this comparison. The oligoarray experiment was repeated for wild-type and *gun5* using both the green and norflurazon-treated samples, with similar results. The *LHCB* probe used for the northern blot analysis was generated from a plasmid containing *LHCB1**2 genomic DNA, and a 25S ribosomal RNA genomic fragment was cloned by PCR and used as a probe for normalization.

Analysis of chlorophyll and tetrapyrrole intermediates

Porphyrin extraction and HPLC analysis were done according to the method described in ref. 25 with slight modification. Leaf material (10–100 mg) was homogenized in 0.67 ml 0.1 M potassium phosphate buffer (pH 7.8) with a HITACHI HG30 homogenizer and centrifuged with ice-cooling. The residue was resuspended in 0.67 ml acetone: 0.1 M NH_4OH (8:2, v/v) and homogenized and centrifuged with ice-cooling. The residue was resuspended again in methanol: 0.1 M NH_4OH (9:1, v/v) and the same procedures described above were repeated. The collected supernatants were mixed and centrifuged before HPLC analysis. Porphyrins were eluted with a linear gradient of solvent B (70% (w/v) acetonitrile and 30% acetone) in solvent A (80% (w/v) methanol and 0.1 M ammonium acetate, pH 5.2) as follows: 0% to 100% over 10 min and 100% solvent B for 20 min. Column eluent was monitored by fluorescence detection ($\lambda_{\text{excitation}}$ 420 nm/ $\lambda_{\text{emission}}$ 595 nm), and porphyrins were identified and quantified using authentic standards. Five μl of standard solutions of MgProtoIX from 10^{-5} to 10^{-11} M were injected into the HPLC; 10^{-11} M solution was the limit for preparing the calibration line in the system used. The detection limit for the HPLC analysis was about 3×10^{-14} g of MgProtoIX. Mg-protoporphyrin IX were purchased from Porphyrin Products and Mg-protoporphyrin IX monomethyl ester was chemically synthesized by treating Mg-protoporphyrin IX with diazomethane in methanol. This procedure gave the mixture of Mg-protoporphyrin IX monomethyl ester and Mg-protoporphyrin IX dimethyl ester, but each of them could readily be identified separately from the retention time in HPLC analysis.

Luciferase activity measurements

Seedlings were ground in liquid nitrogen and the samples extracted in lysis buffer supplied by the Luciferase Assay system (Promega). Protoplasts were collected and lysed in the same lysis buffer. Luciferase activity was detected using a Luminometer (MicroLumat Plus, EG&G Berthold). The different *LHCB1::Luciferase* reporter constructs used in the study are the following: –199::*LUC* contains a promoter fragment of 199 base pairs (bp) from the transcription start of the *LHCB1* gene, *G3M::LUC*, which contains a –199 bp promoter fragment, where the –111 to –33 GT-1 binding site has been mutated, –142::*LUC* contains a promoter fragment of 142 bp from the transcription start of the *LHCB1* gene, and *CUF1M::LUC* contains a –199::*LUC* construct where the –139 to –115 G-box or the CUF1 binding element has been mutated¹⁶. Homozygous lines were established for the different transgenes and for the *gun5* mutation following the crosses.

Received 1 August; accepted 7 October 2002; doi:10.1038/nature01204.

1. Jarvis, P. Intracellular signalling: The chloroplast talks. *Curr. Biology* **11**, 307–310 (2001).
2. Rodermel, S. Pathways of plastid-to-nucleus signalling. *Trends Plant Sci.* **6**, 471–478 (2001).
3. Mochizuki, N., Brusslan, J. A., Larkin, R., Nagatani, A. & Chory, J. *Arabidopsis* genome uncoupled 5 (*gun5*) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proc. Natl Acad. Sci. USA* **98**, 2053–2058 (2001).
4. Surpin, M., Larkin, R. & Chory, J. Signal transduction between the chloroplast and the nucleus. *Plant Cell* **14**, S327–S338 (2002).
5. Mayfield, S. P. & Taylor, W. C. Carotenoid-deficient maize seedlings fail to accumulate light harvesting chlorophyll a/b binding protein (LHCP) mRNA. *Eur. J. Biochem.* **144**, 79–84 (1984).
6. Mayfield, S. P., Nelson, T., Taylor, W. C. & Malkin, R. Carotenoid synthesis and pleiotropic effects in carotenoid-deficient seedlings of maize. *Planta* **169**, 23–32 (1986).
7. Burgess, D. & Taylor, W. Chloroplast photooxidation affects accumulation of cytosolic mRNAs encoding chloroplast proteins in maize. *Planta* **170**, 520–527 (1987).
8. La Rocca, N., Rascio, N., Oster, U. & Rüdiger, W. Amitrole treatment of etiolated barley seedlings leads to deregulation of tetrapyrrole synthesis and to reduced expression of *Lhc* and *RbcS* genes. *Planta* **213**, 101–108 (2001).
9. Susek, R. E., Ausubel, F. M. & Chory, J. Signal transduction mutants of *Arabidopsis* uncouple nuclear *CAB* and *RBCS* gene expression from chloroplast development. *Cell* **74**, 787–799 (1993).
10. Terry, M. J. & Kendrick, R. E. Feedback inhibition of chlorophyll synthesis in the phytochrome chromophore deficient aurea and yellow-green-2 mutants of tomato. *Plant Physiol.* **119**, 143–152 (1999).
11. Oelmüller, et al. Expression of nuclear genes affected by treatments acting on the plastids. *Planta* **168**, 482–492 (1986).

12. Lee, B.-H., Lee, H., Xiong, L. & Zhu, J.-K. A mitochondrial complex I defect impairs cold-regulated nuclear gene expression. *Plant Cell* **14**, 1235–1252 (2002).
13. Crawford, M. S. & Wang, W. Y. Metabolism of magnesium protoporphyrin monomethyl ester in *Chlamydomonas reinhardtii*. *Plant Physiol.* **71**, 303–306 (1983).
14. Sato, R. *et al.* Specific binding of protoporphyrin IX to a membrane-bound 63 kilodalton polypeptide in cucumber cotyledons treated with diphenyl ether-type herbicides. *Plant Physiol.* **96**, 432–437 (1991).
15. Terzaghi, W. B. & Cashmore, A. R. Light regulated transcription. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 211–233 (1995).
16. Andersson, L. A., Teakle, G. R., Martino-Catt, S. J. & Kay, S. A. A circadian clock- and phytochrome-regulated transcription is conferred by a 78 bp cis-acting domain of the *Arabidopsis* CAB2 promoter. *Plant J.* **6**, 457–470 (1994).
17. Kropat, J., Oster, U., Pöppel, G., Rüdiger, W. & Beck, C. F. Chlorophyll precursors are signals of chloroplast origin involved in light induction of nuclear heat-shock genes. *Proc. Natl Acad. Sci. USA* **94**, 14168–14172 (1997).
18. Kropat, J., Oster, U., Rüdiger, W. & Beck, C. F. Chloroplast signalling in the light induction of nuclear HSP70 genes requires the accumulation of chlorophyll precursors and their accessibility to the cytoplasm/nucleus. *Plant J.* **24**, 523–531 (2000).
19. Johanningmeier, U. & Howell, S. H. Regulation of light-harvesting chlorophyll binding protein mRNA accumulation in *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **259**, 13541–13549 (1984).
20. Kittsteiner, U., Brunner, H. & Rüdiger, W. The greening process in cress seedlings. II. Complexing agents and 5-aminolevulinic acid inhibit accumulation of *cab* messenger RNA coding for the light-harvesting chlorophyll a/b protein. *Physiol. Plant* **81**, 190–196 (1991).
21. Papenbrock, J. & Grimm, B. Regulatory network of tetrapyrrole biosynthesis – studies of intracellular signalling involved in metabolic and developmental control of plastids. *Planta* **213**, 667–681 (2001).
22. Forsburg, S. L. & Guarente, L. Communication between mitochondria and the nucleus in regulation of cytochrome genes in the yeast *Saccharomyces cerevisiae*. *Annu. Rev. Cell Biol.* **5**, 153–180 (1989).
23. Damm, B. & Willmitzer, L. *Plant Tissue Culture Manual* (Kluwer Academic, Dordrecht, 1991).
24. Ghassseini, M., Waner, D., Tchier, J., Gribskov, M. & Schroeder, J. I. An integrated *Arabidopsis* annotation database for Affymetrix Genechip® data analysis, and tools for regulatory motif searches. *Trends Plant Sci.* **6**, 448–449 (2001).
25. Mock, H. P. & Grimm, B. Reduction of uroporphyrin decarboxylase by antisense RNA expression affects activities of other enzymes involved in tetrapyrrole biosynthesis and leads to light-dependent necrosis. *Plant Physiol.* **113**, 1101–1112 (1997).
26. Ishikawa, A., Okamoto, H., Iwasaki, Y. & Asahi, T. A deficiency of coproporphyrinogen III oxidase causes lesion formation in *Arabidopsis*. *Plant J.* **27**, 89–99 (2001).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank M. Schmid for help with GeneSpring analysis, H. Chen for promoter analysis, J. Nemhauser for discussions and comments on the manuscript, and L. Barden for assistance with manuscript preparation. We thank Sumitomo Chemical Co. Ltd. for providing us with S23142. This work was supported by a grant from the Department of Energy to J.C. and The Howard Hughes Medical Institute. Å.S. is an EMBO postdoctoral fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C. (e-mail: Chory@salk.edu).

Moesin functions antagonistically to the Rho pathway to maintain epithelial integrity

Olga Speck*, Sarah C. Hughes*, Nicole K. Noren†, Rima M. Kulikauskas* & Richard G. Fehon*

* DCMB Group, Department of Biology, Duke University, Box 91000, Durham, North Carolina 27708, USA, and Department of Cell and Developmental Biology and † the Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA

Two prominent characteristics of epithelial cells, apical-basal polarity and a highly ordered cytoskeleton, depend on the existence of precisely localized protein complexes associated with the apical plasma membrane^{1,2}, and on a separate machinery that regulates the spatial order of actin assembly³. ERM (ezrin, radixin, moesin) proteins have been proposed to link transmembrane proteins to the actin cytoskeleton⁴ in the apical domain,

suggesting a structural role in epithelial cells, and they have been implicated in signalling pathways⁵. Here, we show that the sole *Drosophila* ERM protein Moesin functions to promote cortical actin assembly and apical-basal polarity. As a result, cells lacking Moesin lose epithelial characteristics and adopt invasive migratory behaviour. Our data demonstrate that Moesin facilitates epithelial morphology not by providing an essential structural function, but rather by antagonizing activity of the small GTPase Rho. Thus, Moesin functions in maintaining epithelial integrity by regulating cell-signalling events that affect actin organization and polarity. Furthermore, our results show that there is negative feedback between ERM activation and activity of the Rho pathway.

ERM proteins, highly related members of the larger protein 4.1 superfamily, can exist in an active or inactive conformation⁴. The inactive conformation is brought about by head-to-tail interactions between the FERM (four-point-one, ERM) domain and the carboxy terminus. In this state, ERM actin-binding and membrane association sites are not accessible. A conformational change induced by Rho pathway activation produces the active ERM form. ERM proteins bind actin and cell-type-dependent transmembrane proteins such as CD44, CD43 and several I-CAMs. Genetic tests of hypotheses regarding ERM protein involvement in signalling and cytoskeletal support have been impeded by functional redundancies between these three proteins in mammals. In contrast, *Drosophila* has only one ERM gene, *Moesin*^{6,7}, enabling us to perform such analyses.

To test *Moesin* function *in vivo*, we identified a transposable element insertion in the *Moesin* gene, *P{lacW}l(1)G0323* (Supplementary Fig. 1a). Males hemizygous for this insertion display late larval lethality. This lethality can be rescued by a duplication of region 8B, *Dp(1;Y)619*, as well as by a complementary DNA-based *Moesin*⁺ transgene (Supplementary Fig. 1b). Taken together, these results indicate that *l(1)G0323* disrupts *Moesin* function but does not affect any other essential genes, and is hereafter referred to as *Moe*^{G0323}.

To determine the extent to which *Moesin* function is disrupted by this allele, we examined Moesin protein level and distribution in *Moe*^{G0323} mutants. Moesin protein is not detected in imaginal discs from *Moe*^{G0323} hemizygous larvae by immunofluorescence (Fig. 1d) or immunoblotting (Supplementary Fig. 2, lane 4). Furthermore, animals trans-heterozygous for the *Moe*^{G0323} allele and a deficiency of the 8B region, *Df(1)lz-90b24*, display the same lethal period and have phenotypes identical to *Moe*^{G0323} hemizygous and homozygous mutant larvae (data not shown), establishing that *Moe*^{G0323} is genetically null.

To investigate Moesin function in the morphogenesis of epithelial structures we examined the imaginal discs from *Moe*^{G0323} hemizygous larvae. In wild-type epithelial cells, filamentous actin associates with the cell cortex and accumulates in the apical domain (Fig. 1a, arrow), where it largely coincides with Moesin (Fig. 1b, arrow). In *Moe*^{G0323} cells, F-actin undergoes a marked redistribution: it accumulates in ectopic sites within cells (Fig. 1c, e, filled arrowheads) and becomes depleted from the apical region. To demonstrate that actin is reduced in the apical domain, we expressed a *Moe*⁺ transgene only in the posterior compartment of the mutant disc, generating a 'half-rescued' disc. The rescued cells in the posterior compartment (Fig. 1e, f, open arrowheads) have a wild-type phenotype, and have significantly more filamentous actin in the apical domain than adjacent *Moe*^{G0323} cells (Fig. 1e, arrow).

The imaginal epithelium normally consists of a single layer of tall columnar cells with an obvious apical-basal polarity (Fig. 2a–d). In *Moe*^{G0323} imaginal discs, many cells are situated basal to the epithelial monolayer (Fig. 2e–h, open arrowheads). These cells do not express the junctional marker E-Cadherin (Fig. 2g) or other epithelial polarity markers, including α -Catenin, Coracle, Crumbs or Discs lost (data not shown). These data suggest that in *Moe*^{G0323} organisms, epithelial cells lose intercellular junctions and epithelial

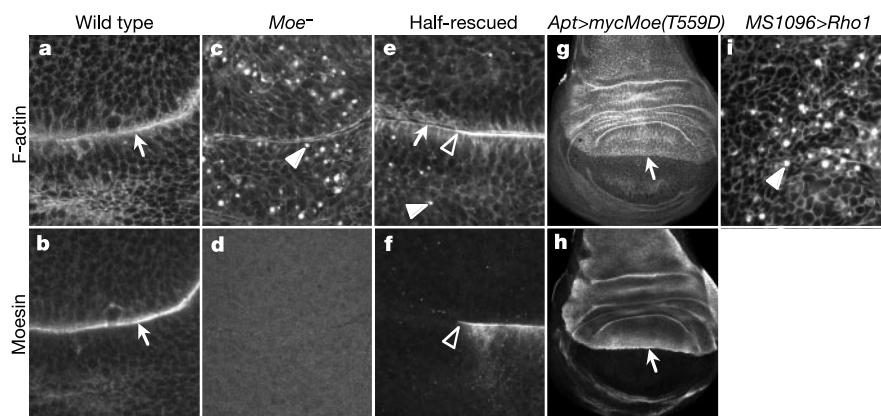


Figure 1 *Moesin* and *Rho1* control filamentous actin. Wing imaginal discs were stained for Moesin (**b, d**) or Myc-tagged Moesin (**f, h**), and filamentous actin (**a, c, e, g, i**). **a, b**, Wild type. Similar to actin, Moesin is enriched in the apical domain (arrows). **c, d**, *Moe*^{G0323} cells have no detectable Moesin staining (**d**) and show accumulations of filamentous actin (**c**, arrowhead). **e–h**, Posterior is to the right, dorsal is up. **e, f**, *Moe*^{G0323}/*Y*; *engrailed-Gal4*^{+/+}; *UAS-mycMoesin*^{+/+}. *UAS-mycMoesin*⁺ rescues the *Moe*^{G0323} actin phenotype in posterior cells that express *engrailed-Gal4*. The

filled arrowhead in **e** indicates an actin accumulation in an anterior cell. Apical actin is reduced in *Moe*^{G0323} cells (arrow) as compared to adjacent rescued cells (**e, f**, open arrowheads). **g, h**, *apterous-Gal4/UAS-mycMoesin(T559D)*. Cortical actin is upregulated in cells that express *UAS-mycMoesin(T559D)*. **i**, *MS1096-Gal4*^{+/+}; *UAS-Rho1*^{+/+}. Actin accumulations (arrowhead) similar to those seen in *Moe*^{G0323} result from *Rho1* overexpression. Magnification: **a–f** and **i** $\times 350$; **g** and **h** $\times 80$.

polarity, and are extruded basally from the epithelium.

To determine whether cells lacking Moesin not only lose epithelial morphology but also adopt migratory behaviour, we marked a subset of cells within the imaginal epithelium by expressing green fluorescent protein (GFP) under control of the region-specific *patched-Gal4* driver. In a wild-type wing imaginal disc, marked cells stay within a continuous stripe along the anterior-posterior boundary (Fig. 3a). In a *Moe*^{G0323} disc, cells expressing GFP under control of *patched-Gal4* are found away from the stripe (Fig. 3d, arrowheads), suggesting that these cells have dispersed. Furthermore, these cells are able to invade adjacent regions of the disc comprising cells that still retain epithelial morphology (Fig. 3c, d, arrow). Thus, *Moe*^{G0323} cells not only lose epithelial markers and

intercellular adhesion, but also become motile and show invasive behaviour, suggesting that cells lacking *Moesin* undergo an epithelial-to-mesenchymal transition.

Rho-induced phosphorylation of a C-terminal threonine has been shown to relieve the inhibitory intramolecular association and generate the active conformation of the ERM molecule^{8–10}. This residue is conserved in *Drosophila* Moesin at amino acid position 559 (ref. 7). To test the role of Thr 559 phosphorylation in Moesin activation, we generated two mutant cDNA constructs: *Moe(T559D)*, a phosphomimetic form, and *Moe(T559A)*, a form that cannot be phosphorylated at this residue. *Moe*^{G0323} hemizygous males are rescued to viability by transgenes expressing *Moe(T559D)*, but not by *Moe(T559A)* (Supplementary Fig. 1b),

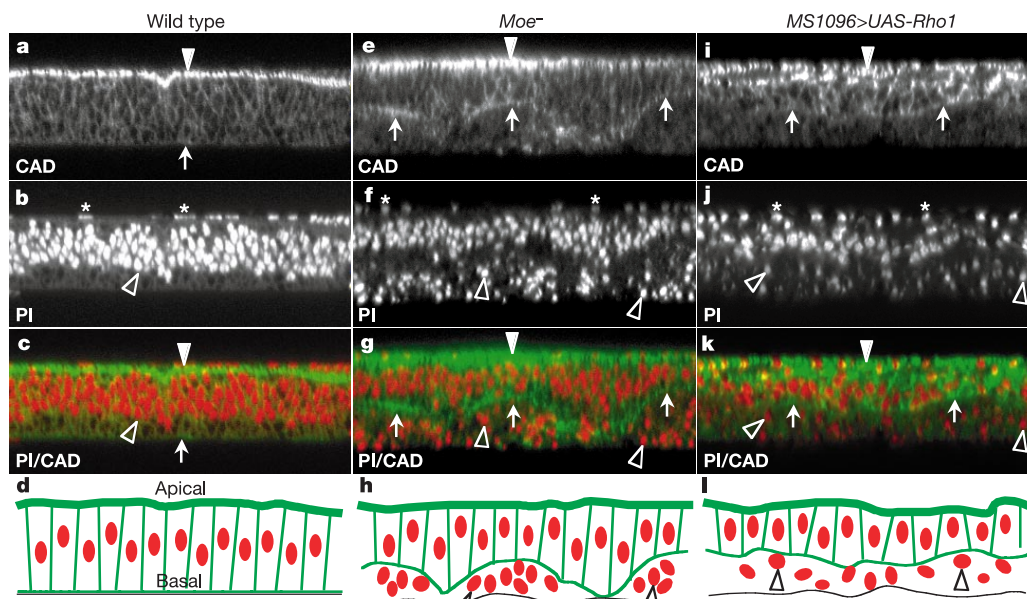


Figure 2 Loss of *Moesin* and overexpression of *Rho1* cause loss of apical-basal polarity. Wing imaginal discs were stained with anti-DE-Cadherin (CAD, green) to highlight the apical domain, and propidium iodide (PI, red) to mark nuclei. Optical sections were taken perpendicular to the epithelium. Filled arrowheads point to apical domains of cells that are in the epithelial monolayer, arrows point to basal domains, and open arrowheads indicate nuclei. **d, h** and **i** are diagrams of merged images in **c, g** and **k**, respectively. **a–d**, Wild

type. All epithelial cells show apical DE-Cadherin staining (**a**) and are organized in a single layer. **e–h**, *Moe*^{G0323}/*Y*. **i–l**, *MS1096-Gal4*^{+/+}; *UAS-Rho1*^{+/+}. Cells lacking apical-basal polarity, as shown by the absence of DE-Cadherin, accumulate basal to the epithelial layer in *Moe*^{G0323} discs as well as in wild-type discs overexpressing *Rho1* (open arrowheads in **f–h** and **j–l**). Asterisks in **b, f** and **j** mark the nuclei of the peripodial membrane. Magnification $\times 230$.

implying that *Moe*(*T559D*) is an active form whereas *Moe*(*T559A*) is inactive. This result contrasts with a recent report¹¹, possibly reflecting differences in how these constructs were tagged.

The phosphomimetic form of ERM proteins has constitutively active properties because it may be locked in the open conformation¹². To examine the effects of ERM activation, we overexpressed *Moe*(*T559D*) under region-specific Gal4 drivers in wild-type imaginal discs. We observed a marked upregulation of cortical F-actin in cells that express *Moe*(*T559D*) (Fig. 1g, h; the arrow points to the boundary of expression). By contrast, little or no effect was seen with *Moe*⁺ and *Moe*(*T559A*) overexpression (data not shown). The accumulation of cortical actin induced by the gain-of-function mutant is complementary to the reduction of apical actin seen in the *Moesin* loss-of-function background, indicating that *Moesin* controls assembly of cortical actin.

The pleiotropic nature of the *Moe*^{G0323} phenotype suggests an involvement in a signalling pathway that controls cell adhesion and motility. Given the crucial function carried out by the Rho family of small GTPases in regulating these processes¹³ and the evidence for Rho regulation of ERM function^{9,14,15} we next examined the relationship between Moesin and Rho signalling *in vivo* by manipulating the genetic dose of *Rho1*, a *Drosophila* *RhoA* homologue, in a *Moe*[−] background. The null *Rho1*^{72R} allele¹⁶ was used to reduce by half maternal and zygotic Rho function. We observed that halving the dose of *Rho1* not only ameliorated the *Moe*^{G0323} disc epithelium organization and actin localization phenotypes (Fig. 4c), but also strongly suppressed *Moe*^{G0323} lethality (Supplementary Table 1). A similar suppression of lethality was observed with other *Rho1* alleles and when the dose of *Drok*, a Rho effector kinase¹⁷, was reduced, and a weaker effect was observed with another downstream effector of *Rho1* signalling, the non-muscle myosin-II heavy chain gene *zipper*^{17,18} (Supplementary Table 1). We determined that the observed suppression was not due to enhanced Moesin protein production or stability (Supplementary Fig 2). Taken together, these results suggest that Moesin functions antagonistically to activity of the Rho pathway in regulating epithelial polarity and integrity.

We examined the ultrastructure of wild-type and *Moe*^{G0323} epithelial cells as well as that of *Moe*^{G0323} cells that are also heterozygous for *Rho1*^{72R}. In transmission electron micrographs, wild-type epithelial cells have apical finger-like microvilli that tend

to congregate near the areas of contact between adjacent cells (Fig. 4d, arrows). In the *Moesin* mutant, the microvilli are replaced by large apical protrusions (Fig. 4e, arrows) that are often of irregular shape. Reduction of Rho activity partially suppresses this *Moe*[−] phenotype, and results in cells that lack the protrusions but still display abnormally large microvillus-like structures (Fig. 4f, arrows). These results suggest that Moesin may have a Rho-independent role in maintaining apical integrity that is not essential for viability.

If Moesin normally functions as an antagonist of Rho pathway activity, then the effects of Rho pathway hyperactivation should be similar to the *Moe* loss-of-function phenotype. To test this hypothesis, we investigated whether *Rho1* overexpression phenotypes in wing imaginal discs mimic those seen in *Moe* loss-of-function mutants. Cells overexpressing *Rho1* mislocalize F-actin (Fig. 1i, arrowhead) and lose epithelial characteristics. As with *Moe*^{G0323} cells, *Rho1*-expressing cells in the blade region of the wing imaginal disc lose junctional markers and drop basally (Fig. 2i–l, open arrowheads). These cells also assume a mesenchymal character, become motile and invade between wild-type epithelial cells (Fig. 3e, f, arrows). Similar, but more disruptive phenotypes were observed when a constitutively active form of *Rho1*, *Rho1*^{V14}, was overexpressed in wild-type imaginal discs (data not shown). Therefore, *Rho1* overexpression strongly resembles *Moe* loss of function,

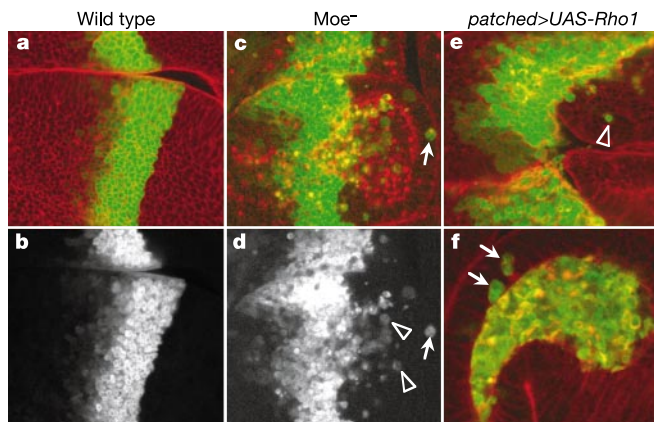


Figure 3 Loss of *Moesin* and overexpression of *Rho1* induce invasive migratory cellular behaviour. Wing imaginal discs were stained with phalloidin (red in **a**, **c**, **e** and **f**). GFP (green in **a**, **c**, **e** and **f**) from **a** and **c** is shown separately in **b** and **d**, respectively. **a**, **b**, *patched-Gal4*⁺; *UAS-GFP*⁺. All GFP-marked cells stay together in the GFP-expressing stripe in a wild-type disc. **c**, **d**, *Moe*^{G0323}/*Y*; *patched-Gal4*⁺; *UAS-GFP*⁺. **e**, **f**, *patched-Gal4*⁺; *UAS-GFP*/*UAS-Rho1*⁺. In contrast to wild type, *Moe*^{G0323} cells and wild-type cells that overexpress *Rho1* disperse from the *patched* stripe (arrowheads in **d** and **e**). Arrows denote migrating cells that have invaded the epithelium. Magnification: **a** and **b** × 180; **c** and **d** × 280; **e** × 320; **f** × 530.

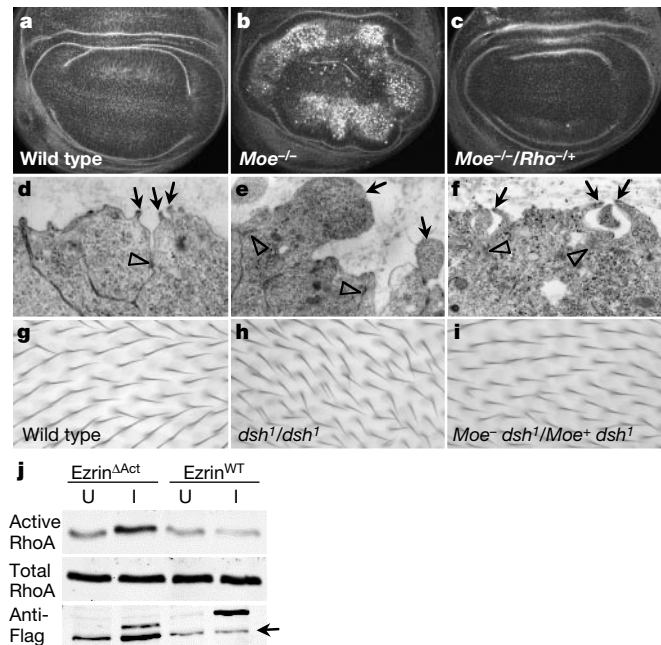


Figure 4 ERM proteins negatively regulate Rho pathway activity in epithelial cells. **a–f**, Wing imaginal discs stained with phalloidin (**a–c**) or processed for transmission electron microscopy (**d–f**; only the apical portions of epithelial cells are shown). **a**, **d**, Wild type; **b**, **e**, *Moe*^{G0323}/*Moe*^{G0323}; **c**, **f**, *Moe*^{G0323}/*Moe*^{G0323}; *Rho1*^{72R}/*+*. The disrupted overall morphology and actin accumulations in *Moe*[−] (**b**), as compared with wild type (**a**), are suppressed by reducing the dose of *Rho1* (**c**). The epithelial ultrastructure in *Moe*[−] is disrupted: large abnormal protrusions (**e**, arrows) replace microvillar projections seen in wild-type cells (**d**, arrows). This phenotype is partially suppressed by reducing the dose of *Rho1*—the apical surface structures resemble wild-type microvilli (**f**, arrows). Arrowheads in **d–f** indicate adherens junctions. **g–i**, Adult wings of indicated genotypes. **j**, Immunoblot of LLC-PK1 cells expressing ezrinΔAct (dominant negative) or wild-type ezrin (ezrin^{WT}). Activated RhoA levels increased in induced (**I**) ezrinΔAct cells relative to uninduced controls (**U**), whereas cells expressing wild-type ezrin showed a slight decrease relative to uninduced controls. Total RhoA in cell lysates served as a loading control. The arrow indicates a nonspecific band recognized by the anti-Flag antibody used to confirm expression of ezrin proteins. Data are representative of multiple independent experiments. Magnification: **a–c** × 100; **d** × 9,500; **e** and **f** × 8,400; **g–i** × 190.

suggesting that *Moesin* and *Rho1* exert opposite effects on the epithelium.

To test further the relationship of *Moesin* with the Rho pathway, we investigated whether a reduction in *Moe* function would suppress a phenotype associated with downregulation of *Rho1*. Previous studies have shown that *Rho1* functions downstream of *dishevelled* (*dsh*) in the planar cell polarity (PCP) pathway that regulates the polarity and number of hairs generated by each wing blade cell^{16,17}. Each wild-type wing cell produces a single hair (Fig. 4g); however, multiple wing hairs result when Rho pathway function is impaired¹⁶. In flies that are mutant for the *dsh*¹ allele, which inactivates the PCP but not the wingless-signalling function of *dishevelled*, we found double-hair-producing cells at a frequency of $6.3 \pm 1.0\%$ (Fig. 4h), consistent with previous results¹⁷. Removal of a single dose of *Moe* suppressed the number of cells with double hairs in *dsh*¹ mutants to $1.2 \pm 0.2\%$ (Fig. 4i), suggesting that Rho pathway function is upregulated in response to reduction in *Moesin* activity.

Our genetic results in *Drosophila* suggest that *Moesin* functions antagonistically to Rho pathway activity. To distinguish between an effect on Rho itself and effects on a downstream component of the pathway, and to extend our results from *Drosophila* to mammalian cells, we examined the effect of reducing ERM function in mammalian LLC-PK1 epithelial cells. We made use of an ERM truncation, ezrin Δ Act, that has dominant-negative properties in flies and reduces function of all three ERM proteins in these cells (as assayed by phospho-ERM staining, a marker for ERM activation¹⁹; data not shown). Expression of ezrin Δ Act resulted in increased levels of Rho activity, whereas expression of wild-type ezrin caused either no effect, or slightly decreased the level of Rho activity (Fig. 4j). These results indicated that ERM proteins negatively regulate Rho pathway function by altering the activation state of Rho itself, rather than a downstream component of the pathway.

The phenotypic analysis presented here clearly demonstrates that *Moesin* function is required to maintain apical-basal polarity, epithelial integrity, and correct actin distribution. This is consistent with *Moesin*'s association with the apical membrane in epithelial cells and the ability to bind filamentous actin^{7,20}. Our results also show that *Moesin* promotes epithelial morphology by functioning antagonistically to Rho pathway activity, rather than by providing structural linkage between the plasma membrane and the cytoskeleton. *Moe*⁻ defects, including lethality, are strongly suppressed when Rho pathway activity is reduced, suggesting that *Moesin* does not have essential structural functions. Furthermore, we have shown that the diverse *Moe*⁻ phenotypes—decreased cortical actin accumulation, loss of epithelial polarity, and acquisition of migratory cell behaviour—can all be attributed to increased Rho pathway activity. Consistent with this model for *Moesin* function, increased Rho activity has been linked to invasive and metastatic phenotypes in mammalian epithelial cells¹³. Because ERM proteins are known downstream effectors of Rho signalling, our results further suggest the existence of a negative feedback loop between ERM proteins and the Rho pathway. This negative feedback mechanism may be important in inhibiting migratory and invasive cellular behaviours that characterize metastatic cells.

Our results with mammalian epithelial cells suggest that the effect of ERM function on Rho signalling occurs upstream to or at the level of Rho itself. Previous studies have detected direct binding between ERM proteins and two known regulators of Rho function, RhoGDI (Rho GDP dissociation inhibitor)²¹ and Dbl GEF (guanine nucleotide exchange factor for Rho)²². Although one of these studies suggested that ERM proteins function to promote Rho activation *in vitro*²¹, the *in vivo* significance of these interactions has not yet been examined. Elucidation of the functional significance of these interactions may be essential to the further understanding of the relationship between ERM proteins and Rho pathway activity. Of note, the proposed negative feedback loop between ERM and Rho

parallels the antagonistic relationship between the ERM-related tumour suppressor Merlin and another Rho-family GTPase Rac²³, and suggests that ERM and Merlin may have similar roles in regulating Rho-family signalling. □

Methods

Drosophila cultures and crosses

*w*¹¹¹⁸ was used for wild-type controls and for the transformation of UAS constructs. The following stocks were given as gifts: *MS109-Gal4* (S. Leever), *Rho1*^{72R} and *UAS-Rho1*⁺ (line 2.1A) (M. Mlodzik), *Rho1*^{E3.10}, *zip*² and *zip*^{Ebr} (D. Kiehart), *Drok*¹ and *Drok*² (L. Luo). All other strains were obtained from the Bloomington *Drosophila* Stock Center at Indiana University. For phenotypic analysis, *Moe*^{G0323} was balanced over *FM7i*, *P[ActGFP]*.

To determine the rescue activity of *UAS-mycMoesin* transgenes, *Moe*^{G0323}/*FM7i*, *P[ActGFP]*; *T80-Gal4/T80-Gal4* females were crossed to the transgenic males or *w*¹¹¹⁸ males as a control. To assess genetic interactions between *Moe*^{G0323} and Rho pathway genes, females of the genotypes *Moe*^{G0323}/*FM7i*, *P[ActGFP]*; *Rho1*^{allele}/*CyO*, *P[ActGFP]*, *Moe*^{G0323} *Drok*^{allele}/*Binscy*, *Moe*^{G0323}/*FM7i*, *P[ActGFP]*; *zip*^{allele}/*CyO* or *Moe*^{G0323}/*FM7i*, *P[ActGFP]* (control) were crossed to *Moe*^{G0323}/*Y*, *Dp(1:Y)619* males. Rescue and lethality suppression were then measured as the percentage of hemi- or homozygous *Moe*^{G0323} offspring that survived to adulthood (calculated from sibling female classes).

Inverse PCR and transgenic constructs

DNA was extracted from heterozygous *l(1)G0323/FM7c* flies and inverse polymerase chain reaction (PCR) with subsequent cycle sequencing of products was carried out according to the protocol of the Berkeley *Drosophila* Genome Project (<http://www.fruitfly.org/about/methods/index.html>). Sequences were obtained from the 5' and the 3' ends of *P[w(+mC)] = lacW/l(1)G0323* and used to determine the precise position of insertion of this transposable element.

The *UAS-mycMoesin*⁺ construct was generated by PCR amplification from a full-length *Moesin* cDNA clone (GenBank accession number L38909 (ref. 7)), subcloning into *StuI*-digested BSSK-Myc shuttle vector²⁴, and subsequent cloning into the pUAS transformation vector²⁵. Mutagenic primers and the QuickChange Site-Directed Mutagenesis Kit (Stratagene) were used to generate T559A and T559D point mutations in the BSSK-mycMoesin⁺ construct. They were then cloned into the pUAS vector. Transformation of all three constructs was performed as previously described²⁶.

Immunoblotting and immunolocalization

Embryos and wandering third instar larvae of the correct genotypes were selected on the basis of GFP fluorescence. Single embryo immunoblots were performed as previously described⁷, except that anti-Moesin antiserum was used at 1:40,000 dilution. The anti- β -tubulin monoclonal antibody E7 (Developmental Studies Hybridoma Bank), developed by M. Klymkowsky, was diluted 1:10,000. For single imaginal disc immunoblots, wing imaginal discs were dissected in PBS and single discs were placed into 10 μ l lysis buffer⁷ and macerated with forceps. We then followed the protocol for single embryo immunoblots. For immunostaining of imaginal discs²⁷, anti-Moesin antibodies (rabbit anti-Moesin; a gift from D. Kiehart) were used at 1:20,000 and anti-DE-Cadherin antibodies (a gift of T. Uemura) were diluted 1:200. Samples were analysed with a Zeiss LSM 410 confocal microscope and images were processed using Adobe Photoshop 5.5.

Wing hair analysis

We prepared wing cuticles as previously described²⁴. All wing hair cells were counted in the same area between longitudinal veins 1 and 2 for at least 8 wings per genotype. The number of multiple wing hair cells was divided by the total number of wing hair cells. Images of wings were taken using the SPOT RT camera and processed in Adobe Photoshop 5.5.

Transmission electron microscopy

Mutant third instar larvae of the correct genotypes were selected on the basis of GFP fluorescence. Wing imaginal discs were dissected in M3 (Sigma) medium and then transferred to 5% glutaraldehyde in 0.05 M phosphate buffer for 30 min. The samples were then post-fixed and treated as previously described^{28,30}. Thin sections were cut on a Reichert-Jung Ultracut and analysed on a Zeiss EM 10A electron microscope.

Cell culture and RhoA activity assays

The LLC-PK1 porcine kidney epithelial cell line was stably transformed with the ecdysone-inducible plasmid pVgRXR (Invitrogen). This line was then transformed with aminoterminal Flag-tagged wild-type mouse ezrin (ezrin^{WT}) or ezrin with the actin-binding domain deleted (amino acids 553–586; ezrin Δ Act), cloned into the pIND vector. Cells were grown in DMEM medium with 10% fetal bovine serum for all experiments. Cells were induced overnight with 1–10 μ M ponasterone in EtOH and used at approximately 70% confluency for Rho activity assays in which the levels of GTP-bound and thus activated Rho were measured as previously described^{29,30}. For immunoblots, mouse monoclonal anti-RhoA (Transduction Laboratories) was used at 1:250 dilution, and rabbit anti-Flag (Sigma) was used at 1:3,000.

Received 19 September; accepted 1 November 2002; doi:10.1038/nature01295.

1. Tepass, U., Tanentzapf, G., Ward, R. & Fehon, R. Epithelial cell polarity and cell junctions in *Drosophila*. *Annu. Rev. Genet.* **35**, 747–784 (2001).
2. Bachmann, A., Schneider, M., Theilenberg, E., Graw, F. & Knust, E. *Drosophila* Stardust is a partner of Crumbs in the control of epithelial cell polarity. *Nature* **414**, 638–643 (2001).

3. Baum, B. & Perrimon, N. Spatial control of the actin cytoskeleton in *Drosophila* epithelial cells. *Nature Cell Biol.* **3**, 883–890 (2001).
4. Bretscher, A., Reczek, D. & Berryman, M. Ezrin: a protein requiring conformational activation to link microfilaments to the plasma membrane in the assembly of cell surface structures. *J. Cell Sci.* **110**, 3011–3018 (1997).
5. Bretscher, A., Edwards, K. & Fehon, R. G. ERM proteins and merlin: integrators at the cell cortex. *Nature Rev. Mol. Cell Biol.* **3**, 586–599 (2002).
6. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
7. McCartney, B. M. & Fehon, R. G. Distinct cellular and subcellular patterns of expression imply distinct functions for the *Drosophila* homologues of moesin and the neurofibromatosis 2 tumour suppressor, merlin. *J. Cell Biol.* **133**, 843–852 (1996).
8. Matsui, T. *et al.* Rho-kinase phosphorylates COOH-terminal threonines of ezrin/radixin/moesin (ERM) proteins and regulates their head-to-tail association. *J. Cell Biol.* **140**, 647–657 (1998).
9. Shaw, R. J., Henry, M., Solomon, F. & Jacks, T. RhoA-dependent phosphorylation and relocalization of ERM proteins into apical membrane/actin protrusions in fibroblasts. *Mol. Biol. Cell* **9**, 403–419 (1998).
10. Simons, P. C., Pietromonaco, S. F., Reczek, D., Bretscher, A. & Elias, L. C-terminal threonine phosphorylation activates ERM proteins to link the cell's cortical lipid bilayer to the cytoskeleton. *Biochem. Biophys. Res. Commun.* **253**, 561–565 (1998).
11. Polesello, C., Delon, I., Valenti, P., Ferrer, P. & Payre, F. Dmoesin controls actin-based cell shape and polarity during *Drosophila melanogaster* oogenesis. *Nature Cell Biol.* **4**, 782–789 (2002).
12. Gautreau, A., Louvard, D. & Arpin, M. Morphogenic effects of ezrin require a phosphorylation-induced transition from oligomers to monomers at the plasma membrane. *J. Cell Biol.* **150**, 193–203 (2000).
13. Schmitz, A. A., Govek, E. E., Bottner, B. & Van Aelst, L. Rho GTPases: signaling, migration, and invasion. *Exp. Cell Res.* **261**, 1–12 (2000).
14. Hirao, M. *et al.* Regulation mechanism of ERM (ezrin/radixin/moesin) protein/plasma membrane association: possible involvement of phosphatidylinositol turnover and Rho-dependent signaling pathway. *J. Cell Biol.* **135**, 37–51 (1996).
15. Mackay, D. J., Esch, F., Furthmayr, H. & Hall, A. Rho- and rac-dependent assembly of focal adhesion complexes and actin filaments in permeabilized fibroblasts: an essential role for ezrin/radixin/moesin proteins. *J. Cell Biol.* **138**, 927–938 (1997).
16. Strutt, D. I., Weber, U. & Mlodzik, M. The role of RhoA in tissue polarity and Frizzled signalling. *Nature* **387**, 292–295 (1997).
17. Winter, C. G. *et al.* *Drosophila* Rho-associated kinase (Drok) links Frizzled-mediated planar cell polarity signaling to the actin cytoskeleton. *Cell* **105**, 81–91 (2001).
18. Halsell, S. R., Chu, B. I. & Kiehart, D. P. Genetic analysis demonstrates a direct link between rho signaling and nonmuscle myosin function during *Drosophila* morphogenesis. *Genetics* **156**, 469 (2000).
19. Hayashi, K., Yonemura, S., Matsui, T. & Tsukita, S. Immunofluorescence detection of ezrin/radixin/moesin (ERM) proteins with their carboxyl-terminal threonine phosphorylated in cultured cells and tissues. *J. Cell Sci.* **112**, 1149–1158 (1999).
20. Edwards, K. A., Demsky, M., Montague, R. A., Weymouth, N. & Kiehart, D. P. GFP-moesin illuminates actin cytoskeleton dynamics in living tissue and demonstrates cell shape changes during morphogenesis in *Drosophila*. *Dev. Biol.* **191**, 103–117 (1997).
21. Takahashi, K. *et al.* Direct interaction of the Rho GDP dissociation inhibitor with ezrin/radixin/moesin initiates the activation of the Rho small G protein. *J. Biol. Chem.* **272**, 23371–23375 (1997).
22. Takahashi, K. *et al.* Interaction of radixin with Rho small G protein GDP/GTP exchange protein Dbl. *Oncogene* **16**, 3279–3284 (1998).
23. Shaw, R. J. *et al.* The Nf2 tumour suppressor, merlin, functions in Rac-dependent signaling. *Dev. Cell* **1**, 63–72 (2001).
24. LaJeunesse, D. R., McCartney, B. M. & Fehon, R. G. Structural analysis of *Drosophila* merlin reveals functional domains important for growth control and subcellular localization. *J. Cell Biol.* **141**, 1589–1599 (1998).
25. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
26. Rebay, I., Fehon, R. G. & Artavanis-Tsakonas, S. Specific truncations of *Drosophila* Notch define dominant activated and dominant negative forms of the receptor. *Cell* **74**, 319–329 (1993).
27. Fehon, R. G., Dawson, I. A. & Artavanis-Tsakonas, S. A *Drosophila* homologue of membrane-skeleton protein 4.1 is associated with septate junctions and is encoded by the coracle gene. *Development* **120**, 545–557 (1994).
28. Prokop, A., Landgraf, M., Rushton, E., Broadie, K. & Bate, M. Presynaptic development at the *Drosophila* neuromuscular junction: assembly and localization of presynaptic active zones. *Neuron* **17**, 617–626 (1996).
29. Ren, X. D., Kiosses, W. B. & Schwartz, M. A. Regulation of the small GTP-binding protein Rho by cell adhesion and the cytoskeleton. *EMBO J.* **18**, 578–585 (1999).
30. Noren, N. K., Liu, B. P., Burridge, K. & Kreft, B. p120 catenin regulates the actin cytoskeleton via Rho family GTPases. *J. Cell Biol.* **150**, 567–580 (2000).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We are grateful to K. Burridge for providing resources for the Rho activation assay; K. Johnson for sharing bench space and expertise; T. Jacks, in whose laboratory the work with mouse ERMs was begun during a sabbatical leave; J. Genova for advice on electron microscopy, and D. Kiehart, D. Lew, I. Rebay, J. Genova and S. Maitra for comments on the manuscript. We would also like to thank H. Gavilan for valuable technical assistance. This work was supported by National Institutes of Health grants to R.G.F. and to K. Burridge.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.G.F.
(e-mail: rfehon@duke.edu).

A cryo-electron microscopic study of ribosome-bound termination factor RF2

Urmila B. S. Rawat*†‡, Andrey V. Zavialov‡§, Jayati Sengupta†, Mikel Valle*†, Robert A. Grassucci*†, Jamie Linde†, Bente Vestergaard||, Måns Ehrenberg§ & Joachim Frank*†¶

* Howard Hughes Medical Institute, Health Research, Inc., Empire State Plaza, Albany NY 12201-0509, USA

† Wadsworth Center, Empire State Plaza, Albany, New York 12201-0509, USA

§ Department of Cell and Molecular Biology, BMC, Uppsala University, Box 596, 5-75124 Uppsala, Sweden

|| Institute of Molecular and Structural Biology, University of Aarhus, DK-8000 Aarhus C, Denmark

¶ Department of Biomedical Sciences, State University of New York at Albany, New York 12222, USA

‡ These authors contributed equally to this work.

Protein synthesis takes place on the ribosome, where genetic information carried by messenger RNA is translated into a sequence of amino acids. This process is terminated when a stop codon moves into the ribosomal decoding centre (DC) and is recognized by a class-1 release factor (RF). RFs have a conserved GGQ amino-acid motif, which is crucial for peptide release and is believed to interact directly with the peptidyl-transferase centre (PTC) of the 50S ribosomal subunit^{1,2}. Another conserved motif of RFs (SPF in RF2) has been proposed to interact directly with stop codons in the DC of the 30S subunit³. The distance between the DC and PTC is ~73 Å. However, in the X-ray structure of RF2, SPF and GGQ are only 23 Å apart⁴, indicating that they cannot be at DC and PTC simultaneously. Here we show that RF2 is in an open conformation when bound to the ribosome, allowing GGQ to reach the PTC while still allowing SPF–stop-codon interaction. The results indicate new interpretations of accuracy in termination, and have implications for how the presence of a stop codon in the DC is signalled to PTC.

A highly efficient *in vitro* system for bacterial protein synthesis was used to generate a release complex (RC) containing ribosomes with a UAA stop codon in the A site and fMet-Phe-Thr-Ile-tRNA^{Ile} in the P site⁵. To study the interaction of the RC with RF2 before and after peptide release, we obtained two types of complex: (1) by reacting wild-type RF2 (RF2(wt)) with RC, in which case the peptide was rapidly removed; (2) by reacting a mutant RF2 (RF2(GAQ)) with RC, in which case the peptide release by RF2 was inhibited. RF2(GAQ) was derived by changing the second glycine residue in the conserved GGQ motif to alanine. This mutation slows down peptide release drastically, but leaves other interactions between factor and ribosome unchanged².

Cryo-electron microscopy (cryo-EM) maps depicting the RC bound to RF2(GAQ) (Fig. 1a) and to RF2(wt) (Fig. 1b) were obtained at resolutions of 10.9 Å and 12.8 Å, respectively. The cryo-maps (Fig. 1a and b) revealed the presence of an electron density attributable to RF2. We see RF2 as a mass in the intersubunit space extending from the DC on the 30S subunit to the PTC on the 50S subunit. There is also extra density in the region of the L7/L12 stalk base of the large subunit that contains the GTPase-associated centre (GAC)⁶. Apart from a small change in the density directed towards the PTC (Fig. 1c and d), the open, tri-lobed shape of RF2 and its position on the ribosome are the same for the mutant (Fig. 1c) and wild-type (Fig. 1d) factors.

The open, tri-lobed shape of the RF2 cryo-EM density before (Fig. 1c) and after (Fig. 1d) peptide release is markedly different from the compact shape of the X-ray structure of RF2 (ref. 4;

Fig. 2a). The fitting of RF2 into the corresponding cryo-EM density was virtually impossible without altering the crystal structure (cross-correlation coefficient (CCC) = 0.3). Instead, we fitted the RF2 structure by allowing domains I and III and superdomain II–IV to rotate as rigid bodies around flexible regions in movements characterized by low-energy barriers⁷. Domains I and III were rotated by $\sim 35^\circ$ and $\sim 75^\circ$, respectively, relative to superdomain II–IV. Protein domain movements in the range $\sim 22^\circ$ to $\sim 150^\circ$ have previously been observed during protein–ligand interactions⁷. We obtained the best fit (CCC = 0.73) by fitting domains I and III into the density directed towards the GAC and the PTC on the 50S subunit, respectively, while placing superdomain II–IV in the thick density strand reaching the DC on the 30S subunit (Fig. 1c and d). The flexible loop region at the tip of domain III containing the GGQ motif was fitted separately.

Domain I of RF2 lies in close proximity to the L7/L12 stalk base, which thus allows its interaction with the amino-terminal domain of protein L11 and with the associated 23S ribosomal RNA segment (Fig. 3a). This ribosomal component forms part of the GAC and is involved in the binding of antibiotics that impede ribosomal G-factor-dependent processes⁸. Interestingly, L11 (ref. 9) and the L7/L12 proteins¹⁰ have been implicated in high-affinity RF–ribosome interactions. It is significant in this regard that binding of RF2(GAQ) to the RC induces a movement in the L11 N-terminal domain by ~ 7 Å, causing it to curl towards the PTC, the only major change in ribosome conformation compared with the RF2-free RC control. This movement is less pronounced for RF2(wt). Indeed, earlier genetic studies suggested correlated changes at the GAC and the PTC during the termination process¹¹.

RF2 domain III, with the conserved GGQ motif at its tip, is

directed towards the PTC on the ribosome, and its distal end lies close to helices 89 and 93 of the 23S rRNA (Fig. 3b), as well as to the CCA end of the P-site tRNA. This is consistent with the proposal that GGQ participates in peptidyl-tRNA hydrolysis^{12,13}. Helix 93 contains nucleotide A2602, situated between the CCA ends of the A- and P-site tRNAs¹⁴. Mutations of A2602 abolish RF-dependent hydrolysis of the ester bond in peptidyl-tRNA (A. Mankin, personal communication). The only conformational difference observed in domain III between RF2(GAQ) and RF2(wt), at the PTC, is a change in the shape of the finger-like density (Fig. 1c and d) that presumably represents the flexible GGQ-loop region. The change might be attributed to the mutation or to a conformational change in RF2 caused by peptide release. The peptide release mediated by RF2(wt) did not alter the position of the P-site or E-site tRNA.

The RF2 superdomain II–IV is directed towards the DC, and comes very close to the stop codon. Helices 34 and 44 of the 16S rRNA and protein S12, known to stabilize codon–anticodon interactions in the DC⁸, are in close proximity to this domain (Fig. 3c). In addition, domain II is near helix 69 of the 23S rRNA, which participates in several intersubunit contacts¹⁵. The SPF motif, a loop region at the tip of domain II, is seen close to the A-site codon (Fig. 3c). This finding strongly supports the earlier conclusion, based on genetic data, that SPF is vital in stop codon recognition³. Direct RF–stop-codon interaction has been suggested by other lines of evidence: (1) RFs and suppressor tRNA compete for the stop codon¹⁶, (2) cross-linking data¹⁷ and (3) hydroxyl radical probing¹⁸.

Because both RFs and tRNAs act at the DC and PTC, RFs have been proposed to mimic tRNAs structurally¹⁹. However, superposition of the two cryo-EM structures of mutant (Fig. 1c) and wild-type (Fig. 1d) RF2 with the X-ray structure of tRNA (aligning parts of the molecules interacting with the DC and PTC separately) shows that the width of the factor is greater, by ~ 10 Å, than that of tRNA, whereas its length along the superdomain II–IV is shorter by ~ 23 Å than the anticodon stem. Accordingly, the overall shape of the ribosome-bound structure of RF2 bears little resemblance to the

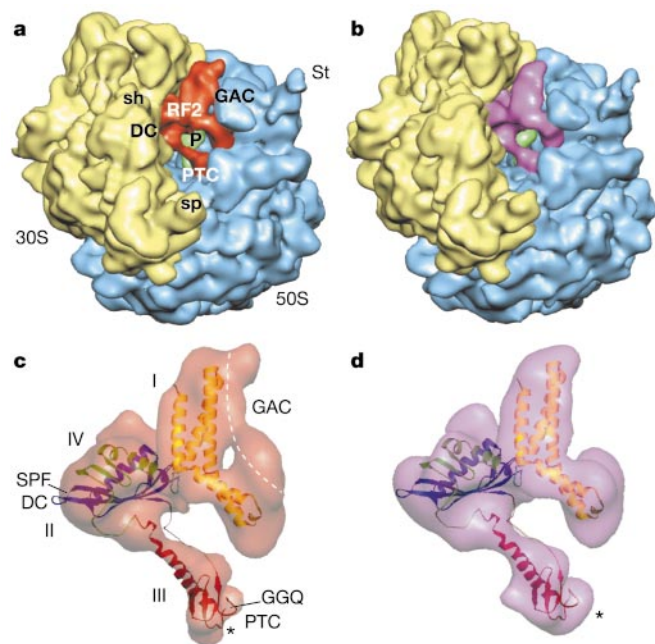


Figure 1 Interaction of RC with RF2. Three-dimensional maps of the RC with RF2(GAQ) mutant (**a**) and RF2(wt) (**b**). Yellow, 30S subunit; blue, 50S subunit. **c, d**, Fitting of modified X-ray structure of RF2 into cryo-density of RF2(GAQ) (**c**) and RF2(wt) (**d**). Ribosome features: sh, shoulder; sp, spur; St, L7/L12 stalk; P, P-site tRNA; DC, decoding centre; PTC, peptidyl-transferase centre; GAC, GTPase-associated centre. Colour coding for RF2: orange, RF2(GAQ); magenta, RF2(wt); gold, domain I; purple, domain II; red, domain III; green, domain IV; magenta, SPF; grey, GGQ. Flexible regions between RF2 domains: cyan, I–II; pale red, II–III; pale green, III–IV. Asterisks in **c** and **d** mark conformational changes in domain III. Density to the right of the dotted line in **c** is due to a conformational change of the ribosome.

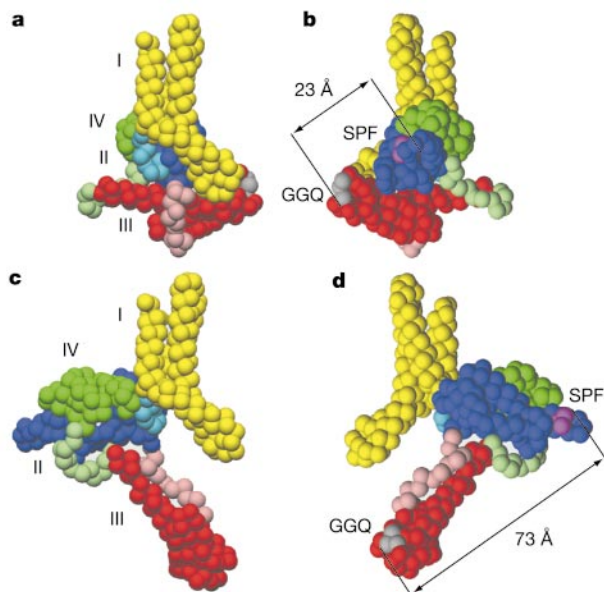


Figure 2 CPK (Corey–Pauling–Koltun) representation of RF2 domain constellations in the two conformations. **a, b**, X-ray structure; **c, d**, cryo-EM structure. The X-ray and cryo-EM structures of RF2 are shown in orientations such that domain I is aligned in each. Colour coding for RF2: gold, domain I; purple, domain II; red, domain III; green, domain IV; magenta, SPF; grey, GGQ. Flexible region between RF2 domains: cyan, I–II; pale red, II–III; pale green, III–IV.

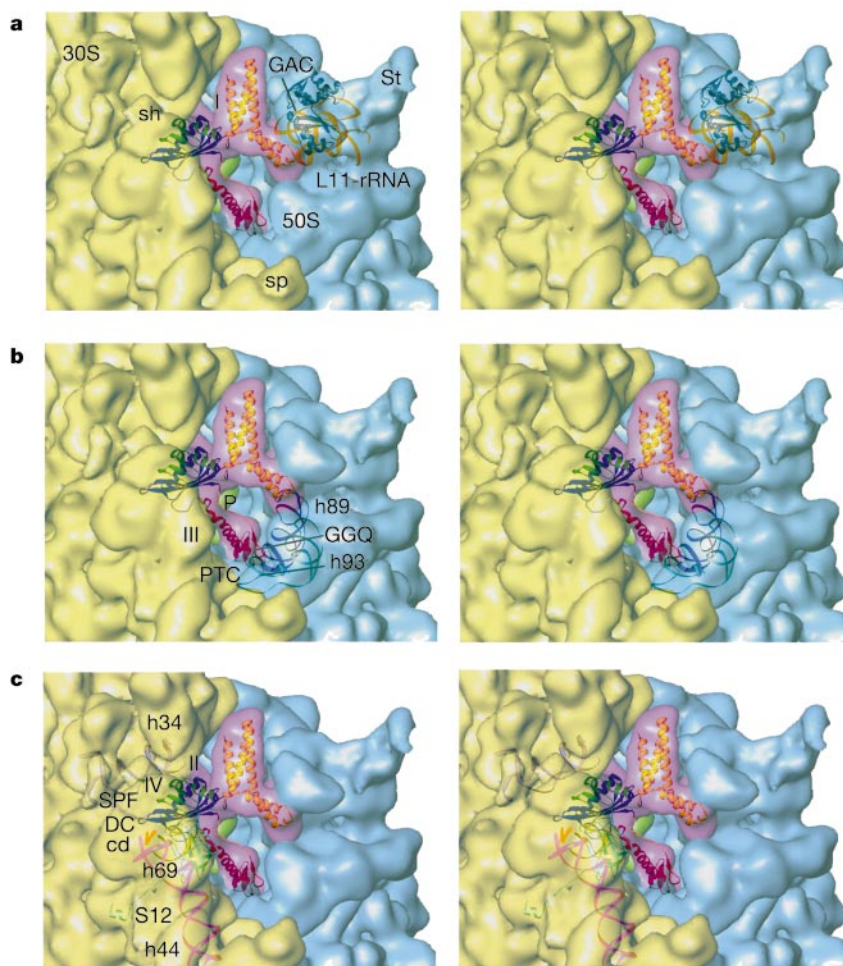


Figure 3 Stereo views showing interaction between functional centres on the RC and RF2 domains. **a**, GTPase-associated centre (GAC): RF2 domain I, protein L11 with 58-nucleotide RNA (L11-rRNA). **b**, Peptidyl-transferase centre (PTC): RF2 domain III with GGQ, helices 89, 93 and P-site tRNA (P). **c**, Decoding centre (DC): RF2 domain II with SPF,

domain IV, helices 44, 34 and 69, protein S12 and A-site codon (cd). Other ribosome features: sh, shoulder; sp, spur; St, L7/L12 stalk. The colour coding for RF2 and the orientation of the complex is the same as in Fig. 1b, with the 30S subunit coloured yellow, the 50S subunit blue, and RF2(wt) magenta; all are semi-transparent.

shape of tRNA. In contrast, our data support the notion of a functional mimicry between RF2 and tRNA, in which SPF and the anticodon of tRNA interact with the mRNA, whereas GGQ and CCA interact with the PTC.

The present study demonstrates that RF2 has an open conformation in its post-termination complex with the ribosome (Fig. 2c and d) that is markedly different from its closed X-ray structure (Fig. 2a and b). In the absence of physical measurements, the conformation of RF2 in solution is unknown. However, the X-ray structures of RF2 determined independently from crystals with different cell parameters are virtually identical⁴, attesting to the stability of this conformation. Furthermore, opening of the closed structure exposes conserved hydrophobic residues at the interface of domain III with superdomain II–IV, which increases the water-accessible surface by 15%. This suggests that a closed rather than an open structure is present also in solution. If so, an attractive new mechanism for the action of RFs could be proposed: RFs would enter the decoding site of the ribosome in a conformation that forms a relatively weak complex with the ribosome. When the RF encounters a sense codon in the DC it dissociates with high probability, but when it encounters its cognate stop codon it attains a fully extended structure that has high affinity for the DC (Figs 1 and 2c, d). Biochemical² and cryo-EM data demonstrate that the extended structures of RF2(GAQ) (Fig. 1a) and RF2(wt) (Fig. 1b) are very strongly bound to the ribosome both before and after the

removal of peptide. The biological importance of strong binding of RFs for ester bond hydrolysis is emphasized by the fact that bacteria have invested in a G protein, RF3, that removes class 1 RFs after peptide release and that they consume one GTP molecule at each termination event^{2,5}.

According to our hypothesis, the presence of a stop codon in the DC is 'signalled' to the PTC by a conformational change of RF2, which increases the distance between the SPF and GGQ and allows the GGQ domain to reach into the PTC in the 50S subunit (Fig. 3b). This proposal has the advantage that the induction of hydrolysis of the ester bond in peptidyl-tRNA is performed by the extended, strongly bound structure of RF2 (Fig. 1a and b), and the precise codon discrimination, which requires a low affinity of RF2 for the ribosome, is performed by a compact, weakly binding form of RF2. □

Methods

RCs were generated as described⁴. Biochemical experiments were performed to study the interactions between RF2(wt) or RF2(GAQ) and RC (complex A). RF2(wt) binding was obtained when complex A (0.13 μ M) was incubated with RF2(wt) (8 μ M) for 2 min at 37 °C, in polymix buffer²⁰ (complex B). Before application to the cryo-grid, an aliquot of complex B was diluted into the same buffer containing 1 μ M RF2(wt), to a final ribosome concentration of 32 nM. In parallel, another aliquot of complex B was withdrawn and subjected to scintillation counting of the [¹⁴C]Ile tetrapeptide released from the P-site tRNA, on the interaction of RF2(wt) with complex A. This yielded a value of ~80% for the occupancy of RF2(wt) in complex B. In conditions similar to those described above for RF2(wt), binding of the RF2(GAQ) mutant to complex A was obtained (complex C). The

occupancy of RF2(GAQ) factor was assumed to be the same as for the wild type, because the factors have similar dissociation constant (K_d) values of 0.7 and 1.8 nM, respectively. Complex A served as the control for complex C, whereas the control for complex B was obtained by incubation of complex A (0.13 μ M) with puromycin (0.4 mM) in the same polymix buffer for 2 min at 37 °C (complex D). Immediately after the incubation of each complex, the samples were stored on ice and then applied to the cryo-grids at a concentration of 32 nM in a cold room at 4 °C. The micrographs were recorded on a Philips FEI (Eindhoven, The Netherlands) Tecnai F20 field emission gun electron microscope at a magnification of 50,000 $\times$ ($\pm$ 2%) and at 200 kV with an electron dose of $\sim 20e^- \text{Å}^{-2}$. Micrographs having Thon rings that showed an absence of astigmatism and drift were selected and scanned on a Zeiss-Imaging scanner (Z/I Imaging Corporation, Huntsville, Alabama, USA) with a sampling step corresponding to a pixel size of 2.82 Å on the object scale. After an automatic particle picking procedure followed by manual and cross-correlation-based selection, totals of 14,965, 18,199, 18,958 and 13,157 ribosome particles were selected for complexes A, B, C and D, respectively. Each data set was subdivided into defocus groups and analysed with the SPIDER system²¹ to obtain refined three-dimensional cryo-density maps. The final resolutions of the CTF-corrected volumes were estimated by using the Fourier shell correlation coefficient with a cut-off value of 0.5, and resolutions of 11.25, 12.8, 10.9 and 14.1 Å were obtained for complexes A, B, C and D, respectively. When assessed by the 3 σ criterion²², the resolutions were in the range 7.4–9.5 Å. The final maps were amplitude-corrected by using low-angle X-ray scattering data²³, and displayed by using IRIS Explorer (Numerical Algorithms Group Inc., Downers Grove, Illinois, USA). SPIDER was used for the computational isolation of the RF2 cryo-EM density. Docking of the X-ray structure of RF2 (pdb code 1GQE), ribosomal proteins and rRNA into the corresponding cryo-densities was performed with O (ref. 24). Individual domains of the RF2 X-ray structure were fitted into the density mass attributable to RF2, and the resulting coordinates were energy-minimized using Insight II to relieve strain from unfavourable steric interactions. Coordinates for the ribosomal proteins and rRNAs were taken from the published atomic structures for the 50S subunit²⁵ (pdb code 1FFK) and the 30S subunit²⁶ (pdb code 1FJF). Atomic coordinates for the A-site codon were obtained from ref. 27 (pdb code 1JGO). The figures were generated using IRIS Explorer, Ribbons²⁸ (University of Alabama at Birmingham) and Insight II LS (Accelrys Inc., San Diego, California, USA).

Received 4 July; accepted 4 October 2002; doi:10.1038/nature01224.

- Kisselev, L. L. & Buckingham, R. H. Translational termination comes of age. *Trends Biochem. Sci.* **25**, 561–566 (2000).
- Zavialov, A. V., Mora, L., Buckingham, R. M. & Ehrenberg, M. Release of peptide promoted by the GGQ motif of class I release factors regulates the GTPase activity of RF3. *Mol. Cell* **10**, 789–798 (2002).
- Ito, K., Ebihara, K. & Nakamura, Y. A tripeptide anticodon deciphers stop codons in messenger RNA. *Nature* **403**, 680–684 (2000).
- Vestergaard, B. *et al.* Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. *Mol. Cell* **8**, 1375–1382 (2001).
- Zavialov, A. V., Buckingham, R. H. & Ehrenberg, M. A posttermination ribosomal complex is the guanine nucleotide exchange factor for peptide release factor RF3. *Cell* **107**, 115–124 (2001).
- Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W. & Ramakrishnan, V. R. A detailed view of the ribosomal active site: the structure of the L11–RNA complex. *Cell* **97**, 491–502 (1999).
- Gerstein, M., Lesk, A. M. & Chothia, C. Structural mechanisms for domain movements in proteins. *Biochemistry* **33**, 6739–6749 (1994).
- Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **108**, 557–572 (2002).
- Tate, W. P., Dognin, M. J., Noah, M., Stöffler-Meilicke, M. & Stöffler, G. The NH₂ terminal domain of *Escherichia coli* ribosomal protein L11. *J. Biol. Chem.* **259**, 7317–7324 (1984).
- Brot, N., Tate, W. P., Caskey, C. T. & Weissbach, H. Requirement for ribosomal proteins L7 and L12 in peptide-chain termination. *Proc. Natl Acad. Sci. USA* **71**, 89–92 (1974).
- Arkov, A. L. & Murgola, E. J. Ribosomal RNAs in translation termination: facts and hypothesis. *Biochemistry (Moscow)* **64**, 1354–1359 (1999).
- Seit-Neibi, A., Frolova, L., Justesen, J. & Kisselev, L. Class-I translation termination factors: invariant GGQ minidomain is essential for release activity and ribosomal binding but not for stop codon recognition. *Nucleic Acids Res.* **29**, 3982–3987 (2001).
- Song, H. *et al.* The crystal structure of human eukaryotic release factor eRF1—mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* **100**, 311–321 (2000).
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
- Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **4**, 897–902 (2001).
- Druege, G. *et al.* Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all the three termination codons in messenger RNA. *Nucleic Acids Res.* **25**, 2254–2258 (1997).
- Tate, W. P., Greuer, B. & Brimacombe, R. Codon recognition in polypeptide chain termination: site directed cross linking of termination codon to *Escherichia coli* release factor 2. *Nucleic Acids Res.* **18**, 6537–6544 (1990).
- Wilson, K. S., Ito, K., Noller, H. F. & Nakamura, Y. Functional sites of interaction between release factor RF1 and the ribosome. *Nature Struct. Biol.* **7**, 866–870 (2000).
- Ito, K., Ebihara, K., Uno, M. & Nakamura, Y. Conserved motifs in prokaryotic and eukaryotic polypeptide release factors: tRNA-protein mimicry hypothesis. *Proc. Natl Acad. Sci. USA* **93**, 5443–5448 (1996).
- Jelenc, P. C. & Kurland, C. G. Nucleoside triphosphate regeneration decreases the frequency of translation errors. *Proc. Natl Acad. Sci. USA* **76**, 3174–3178 (1979).
- Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
- Orlova, E. V. *et al.* Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *J. Mol. Biol.* **271**, 417–437 (1997).
- Gabashvili, I. S. *et al.* Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* **100**, 537–549 (2000).

- Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
- Wimberly, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339 (2000).
- Yusupova, G. Z., Yusupov, M. M., Cate, J. H. & Noller, H. F. The path of messenger RNA through the ribosome. *Cell* **106**, 233–241 (2001).
- Carson, M. Ribbons 2.0. *J. Appl. Cryst.* **24**, 103–106 (1991).

Acknowledgements We thank M. Kjeldgaard (University of Aarhus, Denmark) for pre-releasing the RF2 coordinates, T. Tenson (Tartu University, Estonia) for a critical reading of the manuscript and helpful suggestions, and M. Watters for assistance with the illustrations. This work was supported by the Howard Hughes Medical Institute and by grants from the National Institute of Health (to J.E.) and by the Swedish Foundation for Strategic Research and the Swedish Research Council (to A.V.Z. and M.E.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.E. (e-mail: joachim@wadsworth.org). The coordinates of the modified RF2 structure, obtained by fitting individual domains of RF2 (pdb code 1GQE) into the cryo-EM density of RF2(wt), have been deposited in the Protein Data Bank with accession number 1MI6. The coordinates of the ribosomal components (1G1X, 1G1Y) interacting with the RF2 domains at the GAC, PTC and DC aligned to the cryo-EM map of the 70S *Escherichia coli* ribosome have also been deposited with accession number 1MVR. The cryo-EM maps of the RC (EMD-1006) and its interaction with the RF2(GAQ) mutant (EMD-1008) and the wild type (EMD-1007) have been deposited at the 3D-EM database, EMBL-European Bioinformatics Institute, Cambridge, UK.

Structure of the *Escherichia coli* ribosomal termination complex with release factor 2

Bruno P. Klaholz*, Tillmann Pape*, Andrey V. Zavialov†, Alexander G. Myasnikov*, Elena V. Orlova*‡, Bente Vestergaard§, Måns Ehrenberg† & Marin van Heel*

* Department of Biological Sciences, Imperial College of Science, Technology and Medicine, London SW7 2AY, UK

† Department of Cell and Molecular Biology, BMC, Uppsala University, Box 596, 75124 Uppsala, Sweden

§ Institute of Molecular and Structural Biology, University of Aarhus, 8000 Aarhus C, Denmark

Termination of protein synthesis¹ occurs when the messenger RNA presents a stop codon in the ribosomal aminoacyl (A) site. Class I release factor proteins (RF1 or RF2) are believed to recognize stop codons via tripeptide motifs², leading to release of the completed polypeptide chain from its covalent attachment to transfer RNA in the ribosomal peptidyl (P) site. Class I RFs possess a conserved GGQ amino-acid motif that is thought to be involved directly in protein-transfer-RNA bond hydrolysis^{3,4}. Crystal structures of bacterial and eukaryotic class I RFs have been determined^{5,6}, but the mechanism of stop codon recognition and peptidyl-tRNA hydrolysis remains unclear. Here we present the structure of the *Escherichia coli* ribosome in a post-termination complex with RF2, obtained by single-particle cryo-electron microscopy (cryo-EM). Fitting the known 70S and RF2 structures into the electron density map reveals that RF2 adopts a different conformation on the ribosome when compared with the crystal structure of the isolated protein. The amino-terminal helical domain of RF2 contacts the factor-binding site of the ribosome, the 'SPF' loop of the protein is situated close to the mRNA, and the GGQ-containing domain of RF2

‡ Present address: School of Crystallography, Birkbeck College, University of London, London WC1 7HX, UK

interacts with the peptidyl-transferase centre (PTC). By connecting the ribosomal decoding centre with the PTC, RF2 functionally mimics a tRNA molecule in the A site. Translational termination in eukaryotes is likely to be based on a similar mechanism.

RF2 can be trapped on the ribosome in the absence of GTP and the GTPase RF3, a class II RF, which otherwise accelerates the dissociation of class I RFs after removal of the peptide^{4,7,8}. Incubation of RF2 with ribosomes containing mRNA with a UAA stop codon and a tetrapeptidyl-tRNA leads to a stable ribosomal complex with RF2 in the A site and deacetylated tRNA in the P site. Hydrated single particles of this complex were flash-frozen and analysed by cryo-EM. The structure was determined by angular reconstitution⁹ to 14 Å resolution (as assessed with the 3σ criterion; see Methods and Supplementary Information). Fitting the *Thermus thermophilus* 70S ribosome crystal structure¹⁰ and the *E. coli* RF2 crystal structure⁶ into the electron density map yields a molecular model of the RF2 termination complex with a tRNA molecule in the P site (Fig. 1) and an empty E site. A complete fit of RF2 can be achieved through individual movements of domains 1 and 3 with respect to the functional unit formed by domains 2 and 4. The latter two domains form an elongated shape, which can be fitted unambiguously to the density found close to the decoding site (Fig. 2a). Furthermore, a slight movement of domain 1 (using Gly 121 as a hinge) allows the characteristic shape to fit entirely within

the density extending from the factor-binding site to the 30S head (Fig. 1b). The remaining domain, domain 3, can be fitted to the density extending from the decoding area to the PTC (Fig. 2b). This density in the intersubunit space is sensitive to, for example, 'aggressive' filtering (see Methods), which might be an indication of flexibility in domain 3 or heterogeneity in RF2 occupancy of the sample. The fit was obtained by using hinges in two loop regions located before and after domain 3. These loops are sensitive to protease^{11,12}, have high-temperature factors in the crystal structure⁶ and contain conserved glycine and/or proline residues. The global binding site of RF2 correlates with immunolocalization¹³ as well as site-directed hydroxyl-radical-probing data¹⁴.

Domain 1 bridges the intersubunit space (Fig. 1). The map resolves the three-helix bundle of domain 1 that extends towards helix 33 of the 30S head (Fig. 1b). The other end of domain 1 wraps around the N-terminal domain of protein L11 and is in contact with base A1067 of helix 43 and base A1095 of helix 44 of the 23S RNA. The contact with the 30S head seems to induce a conformational change in the head compared with the 70S crystal structure¹⁰—and compared with the reference structure without RF2 (data not shown)—which includes a movement of helix 33 (30S beak) and a rearrangement of proteins S14, S10, S3 and S5 located around the mRNA entry area (Fig. 1b). The contacts of RF2 with helices 43 and 44 of 23S RNA are in line with data showing that bases A1067 and A1095 are essential for the affinity of RF2 for the ribosome¹⁵. This

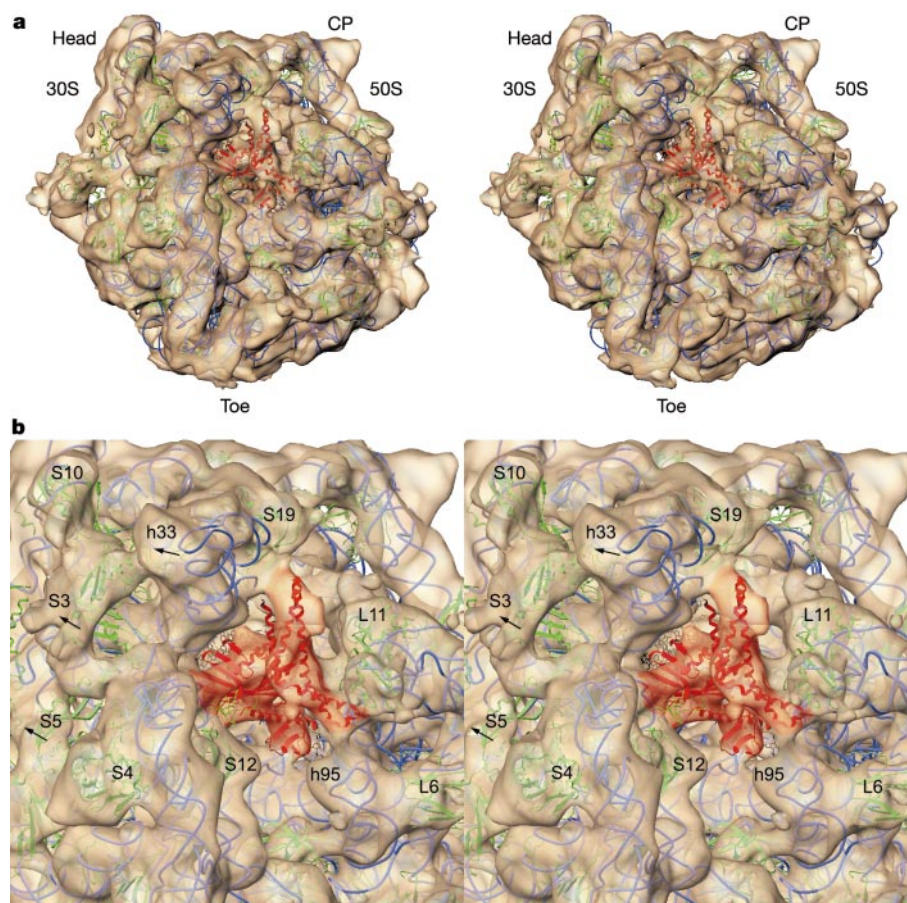


Figure 1 Stereo views of the structure of the post-termination complex revealing the location of release factor RF2. **a**, General view: cryo-EM map (transparent ochre) to which the *T. thermophilus* 70S ribosome crystal structure¹⁰ is fitted as a rigid body (purple, ribosomal RNA; green, proteins; black, P-site tRNA); the fitted *E. coli* RF2 (ref. 6) is shown in red. The subunits of the ribosome (30S and 50S) and major landmarks (head, toe and central protuberance (CP)) are labelled. **b**, Close-up: the density remaining after fitting of

the 70S crystal structure is highlighted in red. The RF2 crystal structure was fitted to that density. RF2 domain 1 is close to L11, helices 43 and 44 of 23S RNA, whereas the three-helix bundle extends towards the 30S head (beak area: helix 33, S3, S10, S14). Note the rearrangements (black arrows) of the 30S beak and proteins S3 and S5 (surrounding the mRNA entry area) with respect to the 70S crystal structure.

area overlaps with the binding site of the antibiotic thiostrepton^{16,17}, suggesting that this drug might be a competitive inhibitor of RF2.

The cryo-EM structure of the RF2 termination complex shows that domain 2/4 of RF2 precisely fits the ribosomal decoding pocket, embedded between helix 18 of 16S RNA and protein S12 on one side, and helices 44 (16S RNA) and 69 (23S RNA) on the other (Fig. 2a). An additional contact can be seen with helix 31 of 16S RNA. RF2 domain 4 is close to helix 18, whereas domain 2 provides all other contacts with the decoding site. The loop in domain 2 that contains the SPF amino-acid motif extends into the mRNA channel. The mRNA is not visible at the current resolution, but its position can be inferred from that in the 70S crystal structure¹⁸, indicating a direct interaction between the SPF loop of RF2 and the stop codon (Fig. 2a). The model suggests that charged residues on the β -sheet surface¹⁹ have only an indirect role in decoding because of their proximity to helices 69 and 44 rather than to the mRNA.

RF2 domain 3 extends from domain 2/4 towards the PTC crossing the intersubunit space (Fig. 2b). The carboxy terminus of helix H7 interacts with helix 69 of 23S RNA and RF2 domain 2, whereas its N-terminus reaches helix 89 of 23S rRNA at the level of the extended loops of protein L16. The β -sheet of domain 3 lies on the rim of helices 71 and 92 (A loop) of 23S RNA and contacts the P-site tRNA-acceptor stem. The flexible GGQ loop at the tip of the β -sheet is not resolved in the map and was not moved with respect to the core of the domain. It has been suggested that the GGQ loop changes conformation when bound to the ribosome⁶ in a manner similar to that for ribosomal protein S5. In its present conformation, the GGQ loop can reach helix 93 of 23S RNA and the 3' end of the P-site tRNA. The P-site tRNA is shifted towards protein L5 compared with its position in the 70S crystal structure; the CCA end

is thus partly withdrawn from the PTC, which might facilitate the access of RF2 domain 3 to the ester bond in peptidyl-tRNA.

The structure of RF2 on the ribosome is different from its crystal structure. Our data reveal that RF2 adopts an open conformation on the ribosome, in contrast to the more compact crystal structure (Fig. 3). On binding of RF2 to the ribosome, recognition of the stop codon is likely to induce a conformational change of RF2 that positions the GGQ loop at the tip of domain 3 in the direct vicinity of residues of the PTC (Fig. 2b). The conformational change resolves the discrepancy between the observed distances in a tRNA molecule and in the crystal structure of RF2. In the crystal structure, the SPF and GGQ motifs are 23 Å apart, a distance very different from the 75 Å that separates the anticodon and the CCA end in a tRNA molecule. The adopted conformation of RF2 allows the release factor to transmit the decoding signal to the PTC, thus mimicking the function of a tRNA molecule. The superposition of the RF2 *in situ* and an A-site tRNA molecule shows that their structures overlap only partly (Fig. 3d), indicating that the 'macro-molecular mimicry' concept^{20,21} for RF2 applies mainly to its function and less to its shape.

Although both prokaryotic and eukaryotic class I release factors are expected to interact similarly with the ribosomal A site, the overall structures found for the isolated factors differ strongly, eRF1 having an open, tri-lobed conformation and RF2 having a compact, bi-lobed structure^{5,6} (Fig. 3). For eRF1, domains 1 and 2 are proposed to represent the codon recognition and peptidyl-tRNA hydrolysis domains, respectively⁵. Biochemical data suggest that eRF1 domain 1 carrying the NIKS amino-acid motif corresponds to RF2 domain 2 carrying the SPF motif^{5,22–24}. The decoding domains of eRF1 and RF2 have a similar shape, despite their different folds,

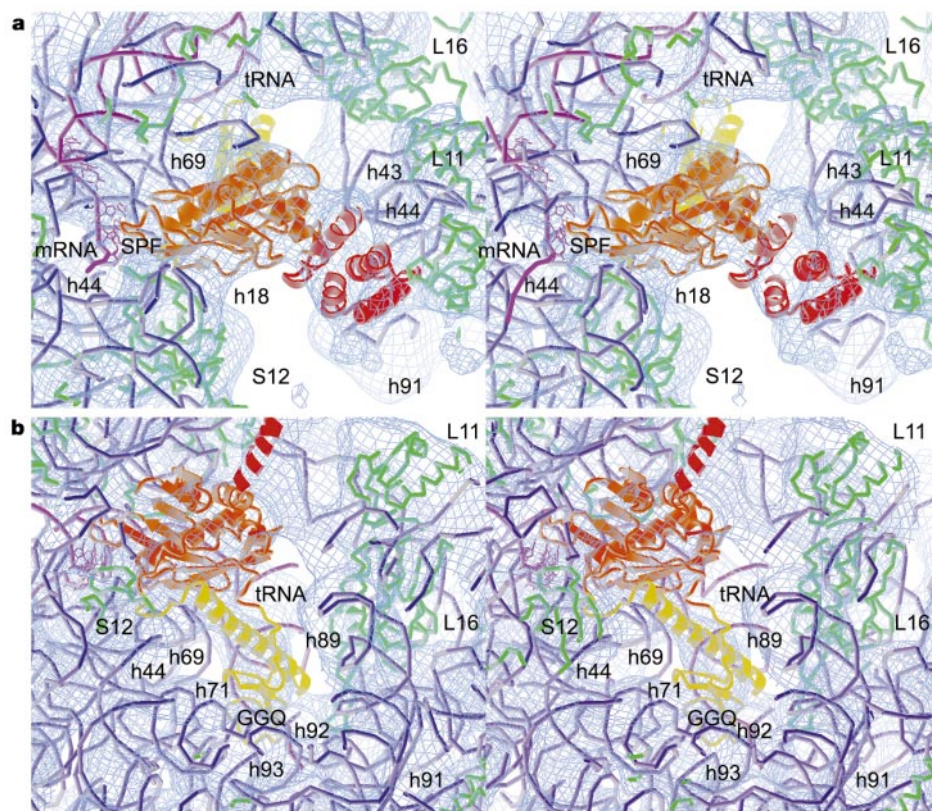


Figure 2 Detailed stereo views of RF2-ribosome interactions at the level of the decoding site and the PTC. Domain 1 of RF2 is shown in red, domain 2/4 in orange, and domain 3 in yellow. Ribosomal RNAs are coloured in blue, proteins in green, and mRNA and the P-site tRNA in magenta. **a**, Decoding centre: domains 2 and 4 form a unit embedded between helix 18 of 16S RNA and protein S12 on one side, and helices 44 of 16S RNA and

69 of 23S RNA on the other. The tip of domain 2 containing the SPF-motif loop can reach the stop codon. **b**, PTC: domain 3 extends from helix 69 of 23S RNA towards the PTC, contacting helices 71, 89 and 92 of 23S RNA, and P-site tRNA. The flexible GGQ loop is close to helix 93 and to the P-site tRNA 3' end.

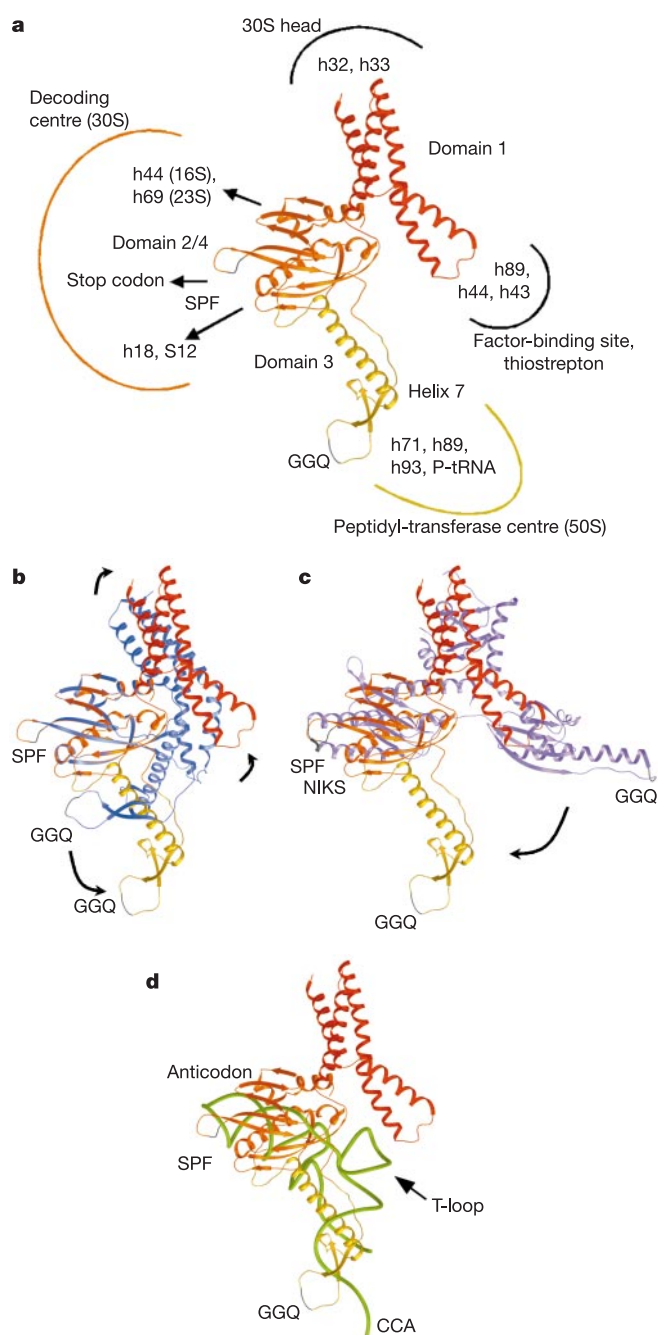


Figure 3 The derived *in situ* RF2 structure and its interaction pattern. The RF2 domains are colour-coded as in Fig. 2. The two functional important sites for decoding (SPF loop) and peptide hydrolysis (GGQ loop) are indicated. **a**, *In situ* RF2 areas involved in transmission of the decoding signal (left) to the PTC (bottom right). The main ribosomal proteins and RNA helices involved are indicated. **b**, Superposition of *in situ* RF2 with the RF2 crystal structure (PDB code 1GQE, shown in blue)⁶ using domain 2/4 as an anchor. The conformational change occurring on binding to the ribosome apparently allows RF2 to contact both ribosomal active sites directly. **c**, Superposition with the eRF1 crystal structure (PDB code 1DT9 (ref. 5), shown in purple) using domain 2/4 of RF2 and the decoding domain 1 of eRF1 as an anchor. Domain 2 of eRF1 is proposed to undergo a rearrangement on binding to the ribosome so that the universally conserved GGQ motifs in this diagram will be in similar positions. **d**, Superposition with A-site tRNA (PDB code 1GIY (ref. 10), shown in green) in its position in the 70S ribosome crystal structure. RF2 seems to mimic a tRNA in functional terms but less so in structural terms. Stereo representations may be viewed in Supplementary Information.

and can be superimposed (Fig. 3c), and the C-terminal domain of eRF1, known to interact with the class II factor eRF3 (ref. 25), can be superimposed with domain 1 of RF2. The folds of eRF1 domain 2 and RF2 domain 3, containing the universally conserved GGQ motif³, are in fact very similar: both consist of a long helix preceded by a long extended loop and a β -sheet. When superimposed on the present *in situ* structure of RF2, the GGQ domain of eRF1 would be required to reorient to reach the PTC (Fig. 3c), adopting a somewhat more closed conformation than in its crystal structure. This rearrangement would be facilitated by the relatively few interactions between domain 2 and the remainder of the molecule, indicating conformational flexibility of domain 2 relative to domains 1 and 3. These considerations indicate that RF2 and eRF1 may have comparable structures on their respective ribosomes and thus termination is likely to be based on similar mechanisms across species. □

Methods

Electron microscopy and image processing

E. coli ribosomes, paused with a UAA stop codon in the A site and a tetrapeptidyl-tRNA in the P site (release complexes, RCs), were prepared by translation *in vitro* and isolated by gel filtration as described⁸. Termination complexes (TCs) were formed in polymix buffer²⁶ by incubating (in 40 μ l) 5.2 pmol RC and 80 pmol RF2 for 2 min at 37 °C. The sample was stored on ice and diluted with polymix buffer containing 0.8 μ M RF2 to a final concentration of 0.011 μ M before being applied to a holey carbon grid at room temperature (20 °C). After excess solution had been blotted off, the grid was rapidly plunged into liquid ethane, leading to ribosome complexes embedded in a thin film of vitrified water spanning the holes on the carbon film. The occupancy of RF2 in TC was ~80%, estimated from the release of tetrapeptide⁸. RCs without addition of RF2 were used as a reference. Electron micrographs were recorded under low-dose conditions at liquid-nitrogen temperature in a Philips CM200 FEG microscope. Images over a whole defocus range (−0.5 to −2.9 μ m) were taken for both samples at a magnification of $\times 50,000$. The quality of the micrographs was checked for minimal drift and astigmatism by optical diffraction. Good micrographs were digitized on an Image Science patchwork densitometer⁹ at a step size of 3 μ m. The digitized images were coarsened by a factor of 4, resulting in a pixel size corresponding to 2.4 Å at the specimen level. Single-molecule images (15,800) were selected interactively from 78 micrographs for studying the termination complex, compared with 3,500 particles from 12 micrographs for the control. Correction of phase contrast transfer function was performed by phase flipping on the raw data. All image processing was performed in the context of the IMAGIC-V software system under Linux and Compaq Tru64 UNIX operating systems⁹. After an initial reference-free alignment (followed by multi-reference alignment, MRA), multivariate statistical analysis (MSA) and automated hierarchical classification of the images, class averages were calculated with an improved signal-to-noise ratio. Subsequently, the Euler angles of these two-dimensional projection images were determined with the angular reconstitution approach by using an anchor set from a previously obtained 70S reconstruction⁹. Three-dimensional (3D) reconstructions were calculated by the exact-filter back-projection algorithm²⁷. Iterative refinements of MRA, MSA and angle refinement were performed until no further improvement could be observed. The resolution of the 3D structures was determined with the Fourier shell correlation (FSC) function (Supplementary Information)²⁷ at a threshold of 3σ (or using a 0.5 cut-off value) with the following results: TC, 14 Å (21 Å); RC, 20 Å (25 Å). For the final representation of the results, the low-frequency and high-frequency components of the 3D map were suppressed by low-pass and high-pass filters corresponding to 1/32 Å and 1/12 Å, respectively. The latter filtering procedure was important for visualizing the map, especially the long helix H7 in domain 3 of RF2. For example, in an earlier representation created with rather sharp cut-off filters, the densities corresponding to RF2 domain 3 were not well defined.

Fitting of crystal structures

Detailed interpretation of the cryo-EM map of the RF2–*E. coli* ribosome complex was largely performed after interactive fitting of the global *T. thermophilus* ribosome crystal structure (pdb codes 1GIX, 1GIY and 1GJQ)^{10,18} into the map with the use of the program O (ref. 28). Subsequently, the *E. coli* RF2 crystal structure (pdb code 1GQE)⁶ was fitted into the remaining electron density. The fitting of RF2 required conformational changes for which hinge areas were first defined, based on an analysis of temperature factors in the crystal structure, on the overall domain organization of RF2, the presence of glycine and/or proline residues, and the known protease cleavage sites as described in the main text. Thus, domain 1, the combined domains 2 and 4, and domain 3 were fitted independently into the available density while maintaining peptide-chain connectivity constraints. Visualization of the density maps in combination with atomic coordinate structures was performed with the O and Amira (TGS Inc.) programs. Figure 3 was generated with the Setor software package²⁹.

Received 4 July; accepted 8 October 2002; doi:10.1038/nature01225.

- Kisselev, L. L. & Buckingham, R. H. Translational termination comes of age. *Trends Biochem. Sci.* **25**, 561–566 (2000).
- Ito, K., Uno, M. & Nakamura, Y. A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature* **403**, 680–684 (2000).
- Frolova, L. Y. *et al.* Mutations in the highly conserved GGQ-motif of class 1 polypeptide release factors

- abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* **5**, 1014–1020 (1999).
4. Zavialov, A. V., Mora, L., Buckingham, R. H. & Ehrenberg, M. Release of peptide promoted by the GGQ-motif of class 1 release factors regulates the GTPase activity of RF3. *Mol. Cell* (in the press).
 5. Song, H. *et al.* The crystal structure of human eukaryotic release factor eRF1-mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* **100**, 311–321 (2000).
 6. Vestergaard, B. *et al.* Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. *Mol. Cell* **8**, 1375–1382 (2001).
 7. Freistoffer, D. V., Pavlov, M. Y., MacDougall, J., Buckingham, R. H. & Ehrenberg, M. Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP dependent manner. *EMBO J.* **16**, 4126–4133 (1997).
 8. Zavialov, A. V., Buckingham, R. H. & Ehrenberg, M. A Posttermination ribosomal complex is the guanine nucleotide exchange factor for peptide release factor RF3. *Cell* **107**, 115–124 (2001).
 9. van Heel, M. *et al.* Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys.* **33**, 307–369 (2000).
 10. Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).
 11. Moffat, J. G. & Tate, W. P. A single proteolytic cleavage in release factor 2 stabilizes ribosome binding and abolishes peptidyl-tRNA hydrolysis activity. *J. Biol. Chem.* **269**, 18899–18903 (1994).
 12. Tin, O. F. *et al.* Proteolytic fragmentation of polypeptide release factor 1 of *Thermus thermophilus* and crystallization of the stable fragments. *Biochimie* **82**, 765–772 (2000).
 13. Kastner, B., Trotman, C. N. & Tate, W. P. Localization of the release factor-2 binding site on 70 S ribosomes by immuno-electron microscopy. *J. Mol. Biol.* **212**, 241–245 (1990).
 14. Wilson, K. S., Ito, K., Noller, H. F. & Nakamura, Y. Functional sites of interaction between release factor RF1 and the ribosome. *Nature Struct. Biol.* **7**, 866–870 (2000).
 15. Xu, W., Pagel, F. T. & Murgola, E. J. Mutations in the GTPase center of *Escherichia coli* 23S rRNA indicate release factor 2-interactive sites. *J. Bacteriol.* **184**, 1200–1203 (2002).
 16. Wimberly, B. T., Guymon, R., McCutcheon, J. P., White, S. W. & Ramakrishnan, V. A detailed view of a ribosomal active site: the structure of the L11–rRNA complex. *Cell* **97**, 491–502 (1999).
 17. Conn, G. L., Draper, D. E., Lattman, E. E. & Gittis, A. G. Crystal structure of a conserved ribosomal protein–rRNA complex. *Science* **284**, 1171–1174 (1999).
 18. Yusupova, G. Z., Yusupov, M. M., Cate, J. H. & Noller, H. F. The path of messenger RNA through the ribosome. *Cell* **106**, 233–241 (2001).
 19. Uno, M., Ito, K. & Nakamura, Y. Polypeptide release at sense and noncognate stop codons by localized charge-exchange alterations in translational release factors. *Proc. Natl Acad. Sci. USA* **99**, 1819–1824 (2002).
 20. Nissen, P. *et al.* Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472 (1995).
 21. Ito, K., Ebihara, K., Uno, M. & Nakamura, Y. Conserved motifs in prokaryotic and eukaryotic polypeptide release factors: tRNA-protein mimicry hypothesis. *Proc. Natl Acad. Sci. USA* **93**, 5443–5448 (1996).
 22. Bertram, G., Bell, H. A., Ritchie, D. W., Fullerton, G. & Stansfield, I. Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. *RNA* **6**, 1236–1247 (2000).
 23. Frolova, L. Y., Seit-Nebi, A. & Kisselev, L. L. Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA* **8**, 129–136 (2002).
 24. Inagaki, Y., Blouin, C., Doolittle, W. F. & Roger, A. J. Convergence and constraint in eukaryotic release factor 1 (eRF1) domain 1: the evolution of stop codon specificity. *Nucleic Acids Res.* **30**, 532–544 (2002).
 25. Merkulova, T. I., Frolova, L. Y., Lazar, M., Camonis, J. & Kisselev, L. L. C-terminal domains of human translation termination factors eRF1 and eRF3 mediate their in vivo interaction. *FEBS Lett.* **443**, 41–47 (1999).
 26. Jelenc, P. C. & Kurland, C. G. Nucleotide triphosphate regeneration decreases the frequency of translation errors. *Proc. Natl Acad. Sci. USA* **76**, 3174–3178 (1979).
 27. Harauz, G. & van Heel, M. Exact filters for general geometry three-dimensional reconstruction. *Optik* **73**, 146–156 (1986).
 28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
 29. Evans, S. V. Setor: Hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).

Supplementary Information accompanies the paper on *Nature's* website (♦ <http://www.nature.com>). A stereo representation of figure 3 and the FSC plot to assess the resolution of the reconstruction are provided.

Acknowledgements We thank M. Kjeldgaard for making the RF2 coordinates available before deposition; E. Murgola for sharing data before publication; E. Morris, R. Finn and R. Matadeen for discussions on image processing; M. Schatz and R. Schmidt for improvements to the IMAGIC software system; G. Willoughby for computational support; and R. Brimacombe for discussions. This work was supported in part by grants from the BBSRC and the EU. B.V. was funded by the NIH. A.V.Z. and M.E. were supported by the Swedish Foundation for Strategic Research and the Swedish Research Council.

Competing interests statement The authors declare competing financial interests: details accompany the paper on *Nature's* website (♦ <http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to M.v.H. (e-mail: m.vanheel@ic.ac.uk). The atomic coordinates of the *E. coli* RF2 structure fitted into the cryo-EM map have been deposited in the Protein Data Bank (e-mail: <http://www.rcsb.org>) with the accession code 1ML5. The cryo-EM map has been deposited in the 3D EM data base (e-mail: http://www.ebi.ac.uk/msd/MSDProjects/IIMS3D_EM.html) with the accession code EMD-1005.

corrigenda

The genome sequence of *Schizosaccharomyces pombe*

V. Wood, R. Gwilliam, M.-A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean, P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M. A. Quail, E. Rabinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schäfer, S. Müller-Auer, C. Gabel, M. Fuchs, A. Düsterhöft, C. Fritzc, E. Holzer, D. Moesti, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dréano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V. A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M. Sánchez, F. del Rey, J. Benito, A. Domínguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, D. Ussery, B. G. Barrell & P. Nurse

Nature **415**, 871–880 (2002).

In this Article, the author Andreas Düsterhöft was mistakenly omitted: his name and affiliation (footnote 6) should have been inserted between M. Fuchs and C. Fritzc in the author list. In addition, the name of L. Cerutti (in the last line of the author list) was misspelled. On p874 in the penultimate sentence of the 'Intergene regions' section, "tandemly oriented genes" should read "divergently oriented genes." □

Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy

Benjamin Schuler, Everett A. Lipman & William A. Eaton

Nature **419**, 743–747 (2002).

The upper limit on the polypeptide reconfiguration time (τ_0) was inadvertently calculated using $(\sigma_{\text{app}} - \sigma_0)^2$, instead of $(\sigma_{\text{app}}^2 - \sigma_0^2)$, as given in the formula in the text (page 745, right column). The correct upper limit is therefore 0.2 ms. This results in a lower limit on the free energy barrier (Δ) of $2k_B T$, corresponding to an activation entropy of $+3k_B$ (page 746, right column), and an upper limit on the pre-exponential factor ($2\pi\tau_0$) of 1 ms. This mistake does not affect any of the conclusions. We thank Taekjip Ha for bringing it to our attention. □

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silke Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakatomachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe
e-mail: h.watanabe@naturejp.com

naturejobs

Out in the cold

The European Commission's Marie Curie Fellowships provide a brilliant way to promote mobility within Europe — they fund postdocs if they move to labs outside their home country. Germany has stumbled on another way to promote mobility, albeit inadvertently: freeze the research budget.

The German government's announcement last month of the freeze is a response to its current economic crisis (see *Nature* **420**, 452; 2002). As a result, as many as 2,000 postdocs could be without funding. That decision is almost sure to cause a brain drain. But the situation begs an obvious question — where do you put 2,000 postdocs?

The answer remains elusive. Even brand new institutes, such as the Institute for Molecular Biotechnology in Vienna, will create only about 200 positions, of which perhaps half will be postdocs (see *Naturejobs* 4–5; 11 April 2002). So if 20 new similarly sized institutions in Europe open this year, the problem would be solved.

Even in the United States, where everything, it seems, is done on a bigger scale, there is no panacea. The closest thing to a quick fix is the Novartis Institute for Biomedical Research, which is occupying existing space in Cambridge, Massachusetts. The drug-discovery facility will employ about 1,000 scientists. But again, in terms of need, the institute falls short. Even if every single position available were for postdocs, it would only absorb about half the fund-seeking Germans.

The solution, if one can really call it that, will involve Germans scattering far and wide, trying to eke out positions wherever growth seems imminent or under way. Or waiting until the government reverses its position and restores the funding.

Paul Smaglik
Naturejobs editor



Contents

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS